

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 17, 2003, 09:52:43 ; Search time 282 Seconds
(Without alignments)
2866.906 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYOHQPOTSNTCTG.....LTENELHROERAPLSKFGD 359

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=xlh
-O=/cgg2_1/USPTO.spool/US09972758/r/natc_17062003_095237_24364/app_query.fasta_1.519
-DB=N-Geneseq,101002 -QPM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=45
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09972758.@CGN.1.1.263.0rnatc_17062003.095237.24364 -NCPU=6 -ICPU=3
-NO_MMP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq,101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1910	100.0	2199	19	AAV82778	Clone bp783_3 isol
2	1910	100.0	2199	24	ABO92015	Human polynucleoti
3	1903	99.6	2178	22	AAH18098	Human cDNA sequenc
4	1902	99.6	1080	24	ABN84013	Human oestrogen do
5	733	38.4	1083	24	ABL90067	Human polynucleoti
6	714	37.4	523	20	AAV86036	EST clone B227. H
7	629	32.9	461	24	ABL63543	Breast cancer rela
8	308.5	16.2	997	22	AAH99667	Human protein enco
9	245	12.8	342	20	AAV87316	EST clone Bp783.
10	214	11.2	1489	23	ABL06283	Drosophila melanog
11	205.5	10.8	3714	23	ABL06282	Drosophila melanog
12	184	9.6	549	21	AAV95217	Cat flea head and
13	174	9.1	1024	23	AAV94493	DNA encoding novel
14	162	8.5	7516	22	AAK51987	Human polynucleoti
15	159.5	8.4	4041	24	ABO54663	Human ovarian anti
16	159.5	8.4	7736	23	AAV65910	DNA encoding novel
17	158.5	8.3	9551	20	AAI22301	DNA encoding a hu
18	156	8.2	6432	23	AAV65903	Human telomerase p
19	155	8.1	2277	19	AAV13836	Homo sapiens mamma
20	155	8.1	2277	19	AAV05372	Human telomerase p
21	153.5	8.0	3579	23	ABL05651	Drosophila melanog
22	153.5	8.0	6105	23	ABL05650	Drosophila melanog
23	152	8.0	2421	24	ABR80600	Bacillus clausii g
24	152	8.0	3540	23	ABL06305	Drosophila melanog
25	151	7.9	3707	23	ABL06276	Drosophila melanog
26	149.5	7.8	8486	22	AAK52971	Human polynucleoti
27	149	7.8	2483	23	ABL07559	Drosophila melanog
28	149	7.8	4483	23	ABL07558	Drosophila melanog
29	149	7.8	18737	23	ABL07492	Drosophila melanog
30	149	7.8	18737	23	ABL07530	Drosophila melanog
31	148.5	7.8	7741	22	AAI60155	Human polynucleoti
32	148.5	7.7	7741	22	AAI60156	Human polynucleoti
33	148	7.7	5943	22	AAV29290	DNA encoding novel
34	147.5	7.7	3733	21	AAI18028	Lung cancer associ
35	147.5	7.7	15606	23	ABL16881	Drosophila melanog
36	147	7.7	7501	22	AAI58369	Human polynucleoti
37	147	7.7	7501	22	AAI58370	Human polynucleoti
38	145	7.6	4301	24	ABR63759	Rat sequence diffe
39	145	7.6	8731	23	ABL10296	Drosophila melanog
40	144	7.5	4714	22	AAV21393	Human cockayne syn
41	143.5	7.5	2159	15	AAV67223	Mouse p53Muc. Mus
42	143	7.5	867	23	AAV84549	DNA encoding novel
43	143	7.5	3399	22	AAV852051	Human polynucleoti
44	142	7.4	2818	23	AAV84552	DNA encoding novel
45	141	7.4	2304	12	AAQ14176	Clone pXRc8 encod

ALIGNMENTS

RESULT 1
AAV82778
ID AAV82778 standard; cDNA: 2199 bp.

XX AAV82778:
DT 25-FEB-1999 (first entry)
XX

DE Clone bp783_3 isolated from human foetal kidney cDNA library.

XX Secreted protein: nutritional activity; immune stimulating; vaccine;
XX suppressing activity; haematopoiesis regulating activity;
XX tissue growth activity; activin; inhibin activity; chemotactaxis;
XX chemokine activity; haemostasis; thrombolytic activity; receptor;
XX ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
XX tumour inhibition; gene therapy; ds.
XX Homo sapiens.
OS

XX MO9842739-A2.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05653.
XX
PR 19-MAR-1998; 98US-0044466.
XX 21-MAR-1997; 97US-0822167.
XX
XX (GEMV) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie BR, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI: 1998-609890/51.
XX P-PSDB; AAW85455.
XX
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, placenta or adult
PT pineal gland cDNA libraries.
XX
PS Claim 1: Page 66-67; 113pp: English.

XX The present sequence encodes a secreted protein. The polynucleotide and
CC secreted protein are predicted to have biological activities which would
CC make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is given.
CC Suggested activities include nutritional activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, actinin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity (no data is
CC given in the specification to support these activities). The
XX polynucleotide is also stated to be useful for gene therapy.

XX SQ Sequence 2199 BP: 552 A; 511 C; 674 G; 462 T; 0 other;

Alignment Scores:
Pred. No.: 6.19e-106 Length: 2199
Score: 1910.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-09-972-758a-2 (1-359) x AAW82778 (1-2199)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
DB 707 ATGCGGAGGAGCCATTCTTGTCAGATATCAACACGAGCTCAAACTAGCACTGTACAGGT 766
QY 21 AlaAlaAlaValGlnGlnGluLeuAsnProGluArgProProGlyAlaGluGluAlaGlyVal 40
DB 767 GCTGCTGCTGCTCCAGGAAGAGCTGAACCTGAGCGGCCCCAGCGGGGAGCGGGGTG 826
QY 41 ProGluGluAspSerArgTTPGlnSerArgAlaPheProGlnLeuGlyGlyAAGProGly 60
DB 827 CCGGAGGAGGACAGTACGTGCGCAATGAGAGCGCTTCCCGCAGTTGGTGGCCGTCCGGG 886
QY 61 ProGluGlyGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
DB 887 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 946
QY 81 GluSerSerCysLeuArgGluGluGlyGluValGlyGlnAsnGlyLysAspSerSerAlaGly 100
DB 947 GAATCTAGCTGCTCTGAGAGAGGCGGAGAAAGGCGCAGAAATGCGAGCACTGCTCGCTGGC 1006
QY 101 GlyAspPheProProProAlaGluValAlaGluProThrProGluAlaGluLeuLeuAlaGln 120
DB 1007 GCGGACTTCCCGCGCGCGGAGAGAGTGGAAACGACGCCCGGAGGCGGAGCTGCTCGCCAG 1066

QY 121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGluGluGlu 140
DB 1067 CCTTGTCATACCTCCAGAGCCCACTAAGTTGGGGGCTCTCCCGAGGGGGCGAAGAGAG 1126
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
DB 1127 TGGGACAGACGACGAGACAGCTGGGCAAGAAAAAATAGAGAGAGCCCGTCAAGAG 1186
QY 161 LysArgHisTrpLysProPyrGlyLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
DB 1187 AAGCGCATTTGGAAACCGTACTACAGCTGACCTGAGAGAGAAAGAAAGTTGACAGAG 1246
QY 181 LysGlnSerLeuArgAlaSerArgAlaAlaAlaLysSerAspSphrSerAspAsp 200
DB 1247 AAGACAGGCGCTTGAGCTTCAAGATCCAGAGCGGAGATGTGCCCAAGGGCCAGCGGTC 1306
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspSphHisAspGlnGluProAspLeu 220
DB 1307 GCGCCCTATACACCCAGCGAGTTCCTCATGATGATCAGACACGAGAGGAGCGCGATCTC 1366
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspSphrSerAspAsp 240
DB 1367 AAAACCGGCGCTGTACTCCAGAGCGGCGCGCCCAAAATCCAGACACACCGATGACAG 1426
QY 241 PheMetGluGluGlyGlyGlyGlyLysArgLysArgLysArgLysArgLysArgLysArg 260
DB 1427 TTCATGGAACAGAGGGGTGAGAGATGGGCGCAGGATGGATGGAGGGAGCGGCGAG 1486
QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
DB 1487 GAATTTCTGACGGGAGCTTCTGAGACCTTACGACCGGTACACACGAGGAGACCTGACAG 1546
QY 281 AsnMetSerLysGlnGluLeuLeuLysGluTyrLeuGluGluGluGluGluGluGluGlu 300
DB 1547 AACATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1606
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAla 320
DB 1607 ATGAGGAGACGACAACACCGCGCTGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1666
QY 321 ArgValArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 340
DB 1667 CGTGTGCGGGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAG 1726
QY 341 ThrGluAsnGluLeuHisArgGlnGlnGluGluGluGluGluGluGluGluGluGluGlu 359
DB 1727 ACCGAGACGAACTGACCGCGGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1783

RESULT 2
ABQ92015 standard: cDNA: 2199 BP.
ID ABQ92015.
AC ABQ92015.
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12.
XX
KW Human: cytostatic; antipneumatic; antiarthritic; vulnery; analgesic;
KW Human: cytostatic; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antilicer; fungicide; antidiabetic; antiallergic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002065394-A1.
XX

ID	AAH18098 standard; cDNA; 2178 bp.
AC	AAH18098;
DT	26-JUN-2001 (first entry)
XX	Human cDNA sequence SEQ ID NO:17952.
DE	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss
KW	Homo sapiens.
OS	Homo sapiens.
XX	EP1074617-A2.
PN	07-FEB-2001.
PD	28-JUL-2000; 2000EP-0116126.
XX	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	(HELI-) HELIX RES. INST.
PA	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Isihii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
PI	WPI; 2001-318749/34.
XX	
XX	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
PS	Claim 8; SEQ ID 17952; 2537bp + CD ROM; English.
XX	
XX	The present invention describes primer sets for synthesizing 5602
CC	full-length cDNAs defined in the specification. Where a primer set
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC	to the complementary strand of a polynucleotide which comprises one of
CC	the 5602 nucleotide sequences defined in the specification, where the
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in
CC	the specification. The primer sets can be used in antisense therapy and
CC	in gene therapy. The primers are useful for synthesizing polynucleotides
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC	represent oligonucleotides, all of which are used in the exemplification
CC	of the present invention.
XX	
XX	Sequence 2178 bp; 525 A; 513 C; 678 G; 462 T; 0 other;
Alignment Scores:	
Pred. No.:	1 61e-105 Length: 2178
Score:	1903.00 Matches: 358
Percent Similarity:	99.72% Conservative: 0
Best Local Similarity:	99.72% Mismatches: 1
Query Match:	99.63% Indels: 0
DB:	22 Gaps: 0

OY	1	MetAlaGluProPheLeuSerGluIyrGlnHISGlnProGlnThrSerGlnSerGly	20
Db	716	ATGGCCGAGCACTTCTGTTCAGAAATATCAACACCAACCTCAAACTAGCACTGACAGG	775
OY	21	AlaAlaAlaValGlnGlnGluLeuGlnProGlnAcrProGlnValAglGluIuArGVal	40
Db	776	GCTCGCTGTCACAGAAAGACTGAACCTTGAGCGCCCCCAGCGCAGAGAGCGG	835
OY	41	ProGlnGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyLysArgProGly	60
Db	836	CCCGAGGAGACAGTAGCTGGCAATCGAGAGGGCTTCCCACTGGGTGGCTCCGGGG	895
OY	61	ProGlnGlyGlnGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro	80
Db	896	CCGAGGGGGAGAGGAGGCTGGAAATCCAACACCACTCCCTTGCAAGCCAGGCTGTCA	955
OY	81	GluSerSerCysLeuArgGluGlyLysGlyGlnAsnGlyAspAspSerSerAlaGly	100
Db	956	GAATCTACTCTGCTGAGAGAGCGCAGAAAGCGCCAAATAGGGAGACACTGTCCTGGAC	101
OY	101	GlyAspPheProProProAlaGluValGluProTrpProGlnValGluLeuLeuAlaGln	120
Db	1016	GAGCACTTCCCGCCCGCCGCAAAATGGAAACCGAGCCCGGAGGCGAGCTGTCCGCCAG	107
OY	121	ProCysHisAspSerGluAlaSerCysLeuGlyAlaProAlaAlaGlyLysGlnGluGln	140
Db	1076	CCTTGCTCATGACTCCGAGGCCAGTAGTTGGGGGCTCTCCGCCGAGGGGGCGAAGGAG	113
OY	141	TrpGlyGlnGlnGlnIuArgGluLeuGlnLysLysLysHisArgArgArgProSerLysLys	160
Db	1136	TGGGGACAGCAGCAGACAGCAGCTGGCGAAGAAACATAGAGAGACGCCGTCCAAAGAG	119
OY	161	LysArgHisTrpLysProTyrTrpCysLeuThrTrpGlnGlyLysLysLysPheAspGln	180
Db	1196	AAGCGCAGTTCGAAACCGTACTACAAAGCTGACCTGGAGAGAGAGAAAGTTTCGACGAG	125
OY	181	LysGlnSerLeuArgAlaSerArgGlyLeuArgAlaGluMerPheAlaLysGlyLysProVal	200
Db	1256	AAACGAGACCTTCGAGCTTCAAGAGATCCGAGCGCGAGTGTGCCAAAGGCCACCGCGTCC	131
OY	201	AlaProTyrAsnThrThrGlnPheLeuLeuLysAspHisAspGlnGluIuProAspLeu	220
Db	1316	GCCCCCTAATACACACGACGAGTCTCTCATGATGATATCAAGACAGAGAGGCGCATCTCC	137
OY	221	LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp	240
Db	1376	AAACCGCGCCTGTACTCCAAAGCGGGCCCGCCCAATCCGAGACACACCGCATGACGAC	143
OY	241	PheMetGlnGluGlnGlyGlnGluAspGlyLysSerAspGlyLysMetGlyLysArgLysArg	260
Db	1436	TTTCATGAGAAAGGGGGTGAGAGGATGGGGGACAGATGGGATGGAGGGAGCGGCAC	149
OY	261	GluPheLeuGlnIuArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln	280
Db	1496	GAGTTTTCGACGCGGACCTTCGAGACGTACGAGCGGTATCCACACGAGAGCGCTCGAG	155
OY	281	AsnMetSerLysGlnGlnGluLeuLysGlyCtyrLeuGlnGluGlnLysCysLeuSerArg	300
Db	1556	AACATGACCAAGCAGAGACTCATCAAGAGATACCTTGGAACCTTGGAAGAGTCCCTCCGCC	161
OY	301	MetGluAspGluAsnAsnArgLeuArgLeuGlnGlnSerLysArgLeuGlnGlyLysAspAla	320
Db	1616	ATGAGAGCAGAGAACACCGGCTGGCGCTCGAGACCAAGCGGCTGGGTGGCAGCAGCGG	167
OY	321	ArgValAlaArgGluLeuGlnGluLeuAlaAspArgLeuArgAlaGluAsnLeuGlnLeuLeu	340
Db	1676	CGTGGCGCGAGCTGGAGCTGGAGTGCAGCGGCTCGCGCGCAGAACCTCCAGCTGCTG	173
OY	341	ThrGluAsnGlnLysLeuHisArgGlnGlnGlnGluArgAlaProLeuSerLysPheLeuLysP	359
Db	1736	ACCGAAGACGAACTGCACCGGACGAGGAGGAGCGGCGGCTTCCAAAGTTTGGAGAC	1792

ABN84013
ID ABN84013 standard; cDNA; 1080 BP.
XX
AC ABN84013:
XX
DT 10-SEP-2002 (first entry)
XX
DE Human oestrogen downregulated gene EDG1 coding sequence.
XX
XX EDG1: oestrogen downregulated gene; tumour suppressor; human:
KM breast cancer; prostate cancer; testicular cancer; ovarian cancer;
KM uterine cancer; colon cancer; chromosome 17q; gene therapy; gene:
KW ss.
XX Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1080
FT /*tag= a
FT /product= "EDG1"
FT /transl_except= "(pos:460..462,aa:Arg)"
FT /transl_except= "(pos:511..513,aa:Thr)"
XX
XX WO200228879-A1.
XX
XX 11-APR-2002.
XX
XX 05-OCT-2001: 2001MO-US31300.
XX
XX 05-OCT-2000: 2000US-238187P.
XX
XX (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX Montano M, Wiltman B:
PI MPI: 2002-519107/55.
DR P-SDB: ABB76495.
XX
XX
XX Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,
PT useful for the prevention, diagnosis and treatment of e.g. breast
PT cancer, testicular cancer, prostate cancer, uterine cancer, cervical
PT cancer, ovarian cancer and colon cancer -
XX
XX
XX Claim 1; Fig 1A-B; 52pp; English.
XX
XX The present sequence is the coding sequence for human oestrogen
CC downregulated gene 1 (EDG1), a tumour suppressor gene that is
CC downregulated by oestrogen in mammary epithelial cells. The gene
CC was identified by yeast two-hybrid screenings for oestrogen
CC receptor-interacting proteins in breast epithelial cells. It
CC was localised to chromosome arm 17q. EDG1 mRNA expression is
CC prevalent in normal mammary epithelial cells and in other human
CC hormone-responsive tissues such as the ovary, prostate and testis.
CC Expression is low in breast cancer epithelial cells. Oestradiol,
CC which induces breast cancer cell growth, has an inhibitory effect
CC on EDG1 mRNA expression in breast cancer cells. Hexamethylene
CC bis-acetamide, an inducer of differentiation and apoptosis,
CC upregulates EDG1 mRNA expression in breast cancer cells. The
CC invention provides EDG1 polynucleotides and polypeptides. In a
CC claimed method, a test sample from an individual suspected of
CC having, or known to have breast, testicular, prostate, uterine,
CC cervical, ovarian or colon cancer is assayed for EDG1 transcript
CC using a polynucleotide that is complementary to the present
CC sequence or by RT-PCR using a primer derived from the present
CC sequence. A decrease in the level of transcript compared to the
CC level in a test sample indicates that the test sample contains or
CC was derived from cancerous cells antibody. A claimed method for
CC decreasing the proliferation of breast, prostate, testicular,
CC ovarian, uterine, cervical or colon cancer cells involves increasing
CC EDG1 protein activity in the cells, either by contacting the cells
CC with EDG1 protein or its fragment or functional equivalent, or with
CC a nucleic acid encoding EDG1 protein, its fragment or functional
CC equivalent.

XX
SQ Sequence 1080 BP; 265 A; 296 C; 376 G; 143 T; 0 other:
Alignment Scores:
Pred. No.: 9.01e-106 Length: 1080
Score: 1902.00 Matches: 357
Percent Similarity: 99.72% Conservative: 1
Best Local Similarity: 99.44% Mismatches: 1
Query Match: 99.58% Indels: 0
DB: 24 Gaps: 0
US-09-972-758a-2 (1-359) x ABN84013 (1-1080)
QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnGlyThrGly 20
DB 1 ATGGCCGAGGCAATCTTGTGCAGAAATATCAACACCACCTCAAACTGACAACTGTACAGGT 60
QY 21 AlaAlaAlaValGlnGlnGlnLeuAsnProGluArgProProGluAlaGlnLeuArgVal 40
DB 61 GCTGCTGCTGTCTCCAGGAAGAGTGAACCTTGACGGCCCCCAGCGCGGAGAGCGGGTG 120
QY 41 ProGluGluAspSerArgTrrpgLInserArgAlaPheProGlnLeuGlyGlyArgProGly 60
DB 121 CCCGAGGAGGACAGTAgGTGGCAATCGAGCGCTTCCCCAGTGGGTGGCCGTCGGGG 180
QY 61 ProGluGluGlySerLeuGlnGlnSerGlnProProProLeuGlnThrGlnAlaGlyPro 80
DB 181 CCGGAGGGGAGGAGGAGGAGCCCTCGAATCCCAACCACTCCCTTGAGAGCCAGGCGCTTCCA 240
QY 81 GluSerSerCysLeuArgGlnGlnGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
DB 241 GAATCTAGCTGCTGTGAGAGAGGCGGAGAAAGGCGCAGAAATGGGAGCAGCTGCTCGCTGGC 300
QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGlnLeuLeuAlaGln 120
DB 301 GCGGACCTTCCCGCGCGCGGAGAAAGTGAACCCAGCCGAGCGCGAGCTGCTCGCCAG 360
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlnGlyGlnGln 140
DB 361 CCTTTCATGACTCCGAGGCGCACTTAAGTTGGGGCTCTGTCGCCAGGAGGGCGAAGAGAG 420
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlnLysLysLysHisArgArgProSerLysLys 160
DB 421 TCGGACAGAGCAGCAGCAGCAGCTGGGGAGAAAAAACAATAGAGAGCCCTTCAGAGAG 480
QY 161 LysArgHisTrpLysProTrrtyrLysLeuThrTrpGlnGlnLysLysPheAspGln 180
DB 481 AAGCGCAATTGGAAACCGTACTCAAGCTGAAGTGGGAAGAGAAAGAAAGTTTCAGAGAG 540
QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGlnIlePheAlaLysGlnGlnProVal 200
DB 541 AAACAGAGCGCTTCGACTTCAGATCCAGCGCAATGTGCCCAAGGGCCAGCGCGTTC 600
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnGlnProAspLeu 220
DB 601 GCCCCTTAACACCCAGCGAGTTCCTCATGGATGATCAGACGAGGAGGCGGGATCTTC 660
QY 221 LysThrGlnLeuThrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240
DB 661 AAACCGGCGCTGTACTCCAAAGCGGGCGCCCAAAATCCGAGCAGCACCGCGATTGACAGC 720
QY 241 PheMetGlnGlnGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
DB 721 TTCATGAGGAAGAGGGGTAGAGAGATGGGGCGAGCGATGGAGATGGAGGGAGCGGCAC 780
QY 261 GlnPheLeuGlnArgAspPheSerGlnThrTrrGluArgTrrHisThrGlnSerLeuGln 280
DB 781 GAATTTCTCAGCGGAGCTTCTCGAGAGCTAGAGACGGGTACACACGAGAGACCTCGCAG 840
QY 281 AsnMetSerLysGlnGlnLeuLysGlnTrrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
DB 841 AACATGAGCAAGCAGGAGCTCATCAAGAGTACTGGAAGTGGAGAGAGTCTTCGCGC 900

QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
 Db 901 ATGGAGGAGCAGAAACACCGCTGCGCTGAGAGCAAGCGCTGGCTGGCAGACGCCG 960
 QY 321 ArgValArgGluLeuGluGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGluLeu 340
 Db 961 CGTGTGCGGAGCGTGGAGCTGAGCTGACCGCTGCGCGCGGAGCACTCCAGCTCTG 1020
 QY 341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
 Db 1021 ACCGAGACGACACTGCAACCGGACAGAGAGACCGCGCTTCCAAAGTTTGCAGAC 1077
 RESULT 5
 ABL90067/c
 ID ABL90067 standard; cDNA; 1083 BP.
 AC ABL90067;
 DT 24-MAY-2002 (first entry)
 XX Human polynucleotide SEQ ID NO 629.
 DE
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KM vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KM neurological disease; infection; human; secreted protein; gene; ss.
 XX Homo sapiens.
 OS
 XX NC0200190304-A2.
 PN 29-NOV-2001.
 PD
 XX 18-MAY-2001; 2001MO-US16450.
 PF
 XX 19-MAY-2000; 2000US-205515P.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Birse CE, Rosen CA;
 PI
 XX WPI: 2002-122018/16.
 DR P-PSDB; ABB89658.
 XX
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -
 PS
 PS Claim 4: SEQ ID NO 629; 2081pp + Sequence Listing: English.
 XX
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPD at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 1083 BP; 301 A; 290 C; 199 G; 286 T; 7 other;

Alignment Scores:
 Pred. No.: 1.16e-35 Length: 1083
 Score: 733.00 Matches: 146
 Percent Similarity: 98.65% Conservative: 0
 Best Local Similarity: 98.65% Mismatches: 2
 Query Match: 38.38% Indels: 1
 DB: 24 Gaps: 0
 US-09-972-758A-2 (1-359) x ABL90067 (1-1083)
 QY 212 AsPHisAspGlnGluGluProAspLeuLysThrGlyLeuTySerLysArgAlaAla 231
 Db 1081 GATCAGCAGCAGAGAGAGCGGACCTCAAAACCGCTTACATCCARCGSCCGCCG- 1023
 QY 232 LysSerAspAspThrSerAspAspAspPheMetGluGluGlyGluGluAspGly 251
 Db 1022 AAATCCGACGACACACGATGACGACTTATGAGAGAGAGGAGGATGGCGGC 963
 QY 252 SerAspGlyMetGlyLysAspGlySerGluPheLeuGluArgAspPheSerGluThr 271
 Db 962 AGCGATGGATGGAGGAGGAGCGACGCGAGCTTCTCGAGCGGAGCTTCTCGAGACCTAC 903
 QY 272 GluArgThrIsthArgLysSerLeuGluAsnMetSerLysGlnGluLeuIleLysGlu 291
 Db 902 GAGCGGTACACACGAGGAGACCTCCAGACATGAGCAAGAGAGACTCATCAAGAGACTAC 843
 QY 292 LeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeu 311
 Db 842 CTGGAACTGAGAAAGTCCCTCTCGCGCATGAGAGAGACAGAAACCGCTGGCGCTGAG 783
 QY 312 SerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuAspArg 331
 Db 782 AGCAAGCGCTGGGTGGCGAGACCGCGCTGCGGAGCTGAGAGCTGGACCGG 723
 QY 332 LeuArgAlaGluAsnLeuGluLeuLeuThrGluAsnGluLeuHisArgGlnGluArg 351
 Db 722 CTGGCGCGCGAGACCTCCAGCTGCTGACCGAGAGACCACTGACCGGAGAGAGCGA 663
 QY 352 AlaProLeuSerLysPheGlyAsp 359
 Db 662 GCGCGCTTCCAAAGTTTGGAGAC 639
 RESULT 6
 AAV86036
 ID AAV86036 standard; cDNA; 523 BP.
 XX
 AC AAV86036;
 XX
 DT 27-APR-1999 (first entry)
 XX
 DE EST clone B227.
 XX
 KM Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KM tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KM chemokinesis; chemokinesis; haemostasis; gene therapy; thrombolytic;
 KM receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 OS
 XX Homo sapiens.
 XX
 PN NC09845435-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98MO-US06954.
 XX
 PR 10-APR-1997; 97US-0835913.
 XX
 XX (GEMV) GENETICS INST INC.
 PA
 PI Asostino Mj, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX
 XX WPI: 1999-070076/06.

XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX
XX
PS Claim 1: Page 97: 633pp; English.
XX
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemokine/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX
SQ Sequence 523 BP; 112 A; 149 C; 153 G; 108 T; 1 other:
XX
Alignment Scores:
Pred. No.: 7-61e-35 Length: 523
Score: 714.00 Matches: 131
Percent Similarity: 99.268 Conservative: 3
Best Local Similarity: 97.048 Mismatches: 1
Query Match: 37.384 Indels: 0
DB: 20 Gaps: 0
US-09-972-758a-2 (1-359) x AAV86036 (1-523)
QY 1 MetaLaGlupProPheUeuSerGluTyrGlnHisGlnProGlnThrsSerAsnCysThrGly 20
DB 104 ATGGCCGAGCCATTCTTGTACAGATATCAACACACACCTCAAACTGTACAGCT 163
QY 21 ALaAlaIValGlnGluGluLeuAsnProGluArgProGluValGluGluVal 40
DB 164 GCTGCTGCTGTCCAGAAAGCTGAACCTGAGCGCCGCCGAGGAGGAGGAGG 223
QY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGluLeuGlyArgProGly 60
DB 224 CCCGAGGAGCAGATGAGTGGCAATGCAGACGCTTCCCGCAGTGGGTGGCCGCGGG 283
QY 61 ProGluGluGluGlySerLeuGluSerGlnProProProLeuGlnThrcAlaCysPro 80
DB 284 CCGGAGGAGGAGGAGGAGCCTGCAATCCCAACCACTCCCTTGAGACCCAGCGCTGCCA 343
QY 81 GluSerSerCysLeuArgGluGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
DB 344 GAATCTAGCTGCTTGAGAGAGGGCGGAGAGGGCCAGAAATGGGAGCACTGCTCGCTGGC 403
QY 101 G1AspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
DB 404 GGGAGCTTCCGCCGCCGAGAGAGTGAAGCGAACGAGCGCCGAGGCGCAGCTGCTGGCCAG 463
QY 121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAla 135
DB 464 CCTTGTCACTAGCTCCGAGCGCAGTAAGATGGGTGCGACGCGCTCT 508
RESULT 7
ABL63543
ID ABL63543 standard; DNA: 461 BP.
XX
XX ABL63543:
XX
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SEQ ID NO:1880.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
OS
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
XX PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-223133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235853P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
PI Sopfel DR, Weaver Z;
XX
XX MPI: 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1: SEQ ID 1880; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I1) of a signature gene set, where (I1)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664

to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour.

XX
XX
S0 Sequence 461 BP; 105 A; 134 C; 138 G; 80 T; 4 other;

Alignment Scores:

Pred. No.:	8,38e-30	Length:	461
Score:	629.00	Matches:	120
Percent Similarity:	93.85%	Conservative:	2
Best Local Similarity:	92.31%	Mismatches:	8
Query Match:	32.93%	Indels:	1
DB:	24	Gaps:	0

US-09-972-758a-2 (1-359) x ABL63543 (1-461)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
|||||
72 ATGCGCGAGCATTTCTGTGAGATATCAACACACGCTTCAACTGACACTGTACAGGT 131
DB 21 AlAlAlAlAlAlGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
|||||
132 GCGTGGCTGTCCAGAGAAAGCTGAACCTGAGCGGCCGCCAGCGNACAGGAGCGCGTTC 191
QY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
|||||
192 CCCGAGAGAGACAGTGTGTGCGATGACAGCCTTCCCGACAGTGGTGGCTCCGCGG 251
DB 61 ProGluGlyGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
|||||
252 CCGAGAGGGGAAAGGAGCGCTGGAAATCCCAACACCTCTTGCAGACCGAGCGCTGTCCA 311
QY 81 GluSerSerCysLeuArgGlnGlyGlyGlyGlnAsnGlnYspAspSerSerAlaGly 100
|||||
312 GAATCTAGCTGCTGTGAGAGAGCGGAGAAAGGCGCAAAATGGGAGCACTGCTGCTGGC 371
DB 101 GlyAspPheProProProAlaGluValGlnProThrProGlnAlaGlnLeuLeuAlaGln 120
|||||
372 GC-GACTTCCCGCCGCCGCGAAGATGAGCCGACGCCCGGAGGCCGACCTGCTTGCACG 430
QY 121 ProCysHisAspSerGlnAlaSerLysLeu 130
|||||
431 CCTGTGATGACTCCGAGGGGAGTAAGTTG 460
DB

RESULT 8
AAH99667/c
ID AAH99667 standard; cDNA; 997 BP.
XX
XX AAH99667;
AC
AC
DT 16-OCT-2001 (first entry)
XX
XX
DE Human protein encoding cDNA sequence SEQ ID NO:502.
XX
XX Human: cancer; ulcer; HIV infection; human immunodeficiency virus;
KM antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KM antibacterial; endocrine; cardiac; central nervous system; virulide;
KM anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KM antiaggregant; haemostatic; vulnary; antidiabetic; osteopathic; eczema;
KM dermatological; antiallergic; antidiabetic; antidiabetic; cytosolic;
KM neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KM immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KM antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KM cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;

genetic disease; haematopoietic disorder; platelet disorder; asthma;
KM thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KM allergic rhinitis; diabetes; multiple sclerosis; depression;
KM Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KM neurological disorder; ss.

OS
XX Homo sapiens.
XX
XX MO200153455-A2.
XX
XX 26-JUL-2001.
XX
XX
XX 22-DEC-2000; 2000MO-US35017.
XX
XX
XX 23-DEC-1999; 99US-0471275.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT.
XX
XX WPI: 2001-457603/49.
XX P-PDB: AAM25726.
XX
XX
XX Claim 1: Page 566-567; 1217pp; English.

AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antidiabetic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virulide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;
CC antidiabetic; osteopathic; dermatological; antiallergic; antidiabetic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmune, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.

XX
XX
S0 Sequence 997 BP; 184 A; 323 C; 266 G; 224 T; 0 other;

Alignment Scores:

Pred. No.:	3.08e-10	Length:	997
Score:	308.50 <td>Matches:</td> <td>70</td>	Matches:	70
Percent Similarity:	60.26% <td>Conservative:</td> <td>21</td>	Conservative:	21
Best Local Similarity:	46.36% <td>Mismatches:</td> <td>43</td>	Mismatches:	43
Query Match:	16.15% <td>Indels:</td> <td>17</td>	Indels:	17
DB:	22	Gaps:	3

US-09-972-758a-2 (1-359) x AAH99667 (1-997)

QY 202 ProTyrAsnThrThcInPheLeuMetAspHisAspGlnGluGluPro-----Asp 219
|||||
980 CCTTACACACACACCGAGTTCTCGATGATGACAGGAGCCGAGAGGCCCAACTTGGAT 921
DB 220 LeuLysThcGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspThrSerAspAsp 239
|||||
920 GTGCCCATAGGAGATCTCCACCCAGGTTCCACTGGGAGAGT----- 879
QY 240 AspPheMetGluGluGlyGlyGluGluGluGluGluGluGluGluGluGluGluGluGlu 259

Db	878	-----GAGCGCGGGACAGTGAATGGCGGGCCCGACGCCAC	843
Qy	260	SerGluPheLeuGlnArgAspPheSerGluThrTyGluArgTyriHisThrGluSerLeu	279
Db	842	GGTGAGTTCAGAGCGAGAGACTTCTGTGAGACTTACGAACGCTTCCACACCGAGACCTG	783
Qy	280	GlnAsnMetSerLysGlnGluLeuIleLysGluTyrlrLeuGluLeuLysCysLeuSer	299
Db	782	CAGGCGCCACACACAGAGAGCTGTGCGACACTACCTGGAGGTGGAGAACCCTGCTCG	723
Qy	300	ArgMetGluAspGluAsnAsnArgGluArgLeuGluSerLysArgGluGlyGluAspAsp	319
Db	722	CAGCGCGGAGAGGAGACTAGCAGGCTGCAGCAGCTGCACGCGTGCACCGCCAGCAGTCC	663
Qy	320	AlaArg---ValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGln	338
Db	662	TGCGCGCCAGGTGGAGAGCTGCTGCGCAGAGTCCAGAGGCTCCGACCGCAAAACGACGGC	603
Qy	339	LeuLeuThrGluAsnGluLeuHisArgGlnGln	349
Db	602	CTTGTGTCAGAGAACCACTGTGGAACCGACAG	570
RESULT 9			
AAV87316			
ID	AAV87316	standard: cDNA: 342 BP.	
XX	AAV87316:		
AC			
XX	27-APR-1999	(first entry)	
DT			
XX	EST clone BP783.		
DE			
XX	Expressed sequence tag; secreted protein; haematopoiesis regulator;		
KW	tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;		
KM	chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;		
KM	receptor; ligand; anti-inflammatory; tumour inhibitor; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	W09845435-A2.		
PD	15-OCT-1998.		
XX			
PF	10-APR-1998:	98WO-US06954.	
XX			
PR	10-APR-1997:	97US-0835913.	
XX			
PA	(GENY) GENETICS INST INC.		
XX			
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D:		
PI	Racie LA, Spaulding V, Treacy M:		
XX			
DR	WPI: 1999-070076/06.		
XX			
PT	New polynucleotides encoding human secreted proteins - derived from		
PT	e.g. human blood, kidney, foetal lung, placenta, testes, brain,		
PT	ovary, pituitary, retina and colon cDNA libraries		
XX			
PS	Claim 1: Page 526: 633pp: English.		
XX			
CC	This sequence represents an expressed sequence tag (EST), and is a		
CC	polynucleotide of the invention. The polynucleotides of the invention are		
CC	all secreted EST sequences isolated from a variety of human tissue		
CC	sources. The EST sequences and proteins encoded by them are predicted to		
CC	have useful biological activities and proteins which would make them suitable for		
CC	treating, preventing or ameliorating medical conditions in humans and		
CC	animals, although no supporting data is given. Suggested activities		
CC	include nutritional activity, immune stimulating or suppressing activity,		
CC	haematopoiesis regulating activity, tissue growth activity,		
CC	activin/inhibin activity, chemotactic/chemokineic activity, haemostatic		
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory		
CC	activity, cadherin/tumour invasion suppressor activity, tumour inhibitor		

CC	activity: The EST sequences are also stated to be useful for gene therapy.
XX	
SQ	Sequence 342 BP; 90 A; 79 C; 83 G; 89 T; 1 other:
	Alignment Scores:
Pred. No.:	6.61e-07 Length: 342
Score:	245.00 Matches: 46
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	12.83% Indels: 0
DB:	20 Gaps: 0
US-09-972-758A-2 (1-359), x AAV87316 (1-342)	
OY	1 MetAlaGluPProPhelLeuSerGIutyrGlnHisGInProGlnThrSerAncysTrhgly 20
Dd	196 ATGGCGGAGCCATTCTTGTGCAGAATATCAACACCACGCCCTCAAAGTAGCAACTGTACAGGT 255
OY	21 AlaAlaIaValGlInGluGluLeuAnSProGluArGProProGlyAlaGluGluArGval 40
Dd	256 GCTGTCGCTGCCAGGAAGAGCTGAACCTTGAGCGGCCCCCGAGGCGGGAGGCGGGTG 315
OY	41 ProGluGluASPSeArG 46
Dd	316 CCCGAGGAGCAGCATGAG 333
RESULT 10	
ABL06283	
ID	ABL06283 standard: cDNA: 1489 BP.
XX	
AC	ABL06283;
XX	
DT	26-MAR-2002 (first entry)
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 13331.
XX	
KW	Drosophila: developmental biology; cell signalling; insecticide;
KM	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NV.
XX	
P1	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI: 2001-656860/75.
XX	
DR	P-PSDB: ABB62180.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PP	
XX	
PS	Claim 1: SEQ ID NO 13331; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid/detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI6176-ABU30511), expressed DNA sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB7737-ABR772C).
CC	The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 1489 BP; 434 A; 363 C; 383 G; 309 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Mismatches:	Indels:	Gaps:
Score: 0.000216	1489	73	46	69	10
Percent Similarity: 42.20%					
Best Local Similarity: 25.89%					
Query Match: 11.20%					

US-09-972-758A-2 (1-359) x ABL06283 (1-1489)

```

OY 85 LeuAtgGluGlyGluLysGlyGlnAsnGlyAsp-----AspSerSerAla 99
DB 98 ATGGCTGAAGCTGTAAATAATGAAGTGGCTCCCAACAGACCTTTGGATAGCGGCGGA 157
OY 100 GlyGlyAspPheProProProAlaGluGluProThrProGluAlaGluLeuAla 119
DB 158 GGAGGGGGC----- 166
OY 120 GlnProCysHisAspSerCysLeuLysLeuGlyAlaProAlaAlaGlyGlyGlu 139
DB 167 -----GCCAGTGGCGGAGCGGCTGCTGAGAGCGCGCTCG 205
OY 140 GluTrpGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 159
DB 206 GGAATGCCGAGAGAGAAACATCGCGGGGAAAAAATCC--AAATGCAAGCCCAAGAA 262
OY 160 LysLysAlaGlnHisTrpLysProTyrTrpLysLeuThrTrpGlu-----GluLys 175
DB 263 ACCAAGAACCAATAT--CCGCAGTGAAGAACATGATATGTCACCGCTGCAGAGCAACG 319
OY 176 LysLysPheAspGluLysGlnSerLeuArgAlaSerArgGlyLeuAlaGluMetPheAla 195
DB 320 CTGGAGGGGATCAACAGACAGATAGCAGAACCAAGCTGGTTCGTCGGCTCTTG--- 376
OY 196 LysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspHisAspGln 215
DB 377 -----CTCGTCCGACACAGAACCGCTTCCTAATGAGAACACACTGTCG 424
OY 216 GluGluProAspLeuLysThrGlyLeuTyrSerLysAlaAlaAlaLysSerAsp 235
DB 425 GAG-----CTGCACAAACATGAC 442
OY 236 ThrSerAspAspPheMetGluGluGlyGlnGluGlnGluGlnGluGlnGluGln 255
DB 443 TCCGACGACACTGTTT-----GGCTCCAGACCGAA----- 475
OY 256 GlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHis 275
DB 476 -----GATCAGGTGCTCTTCTTCCAAAGAGTTCTCGATGTCAGACGAGCGCG 529
OY 276 ThrGluSerLeuGlnAsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGln 295
DB 530 CTCGAAAGCTTTGGAGACGATGAGCAAGCAGCTCATCCAGAGTGCATCAATGCA 589
OY 296 LysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeu 315
DB 590 GATGATATTTCGAAGGCCCAACATATCCAAAGAGTT-----GGAGCCAAATTA 640
OY 316 GlyGlyAspAspAlaArgValArgGluLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 335
DB 641 CGAGCGAGATAGATGATGCGCAACTATCAGCTGAGATCAATTTCTCGGACCCAC 700
OY 336 AsnLeuGlnLeuLeuThrGlnAsnGlu-LeuHisArgGlnGlnGlnGlnGlnGln 355
DB 701 CTCTTGGCGACAGTGTTCGCGACAGCTGGACCAAGCGGACAGCGGCGCGAGCTTCC 760
OY 355 rLys 356
:::

```

DB 761 CCGC 764

RESULT 11

ID ABL06282/C

XX ABL06282;

AC 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13328.

KW Drosophila: developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PA 11-JUL-2000; 2000US-0614150.

PI (PEKE) PE CORP NY.

DR Venter JC, Adams M, Li PWD, Myers EW;

PI WPI: 2001-656860/75.

PT P-PSDB: ABB62179.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PS interactions -

CC Claim 1; SEQ ID NO 13328; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.

XX SQ Sequence 3714 BP; 971 A; 891 C; 852 G; 1000 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Mismatches:	Indels:	Gaps:
Score: 0.00178	3714	69	42	79	10
Percent Similarity: 205.50%					
Best Local Similarity: 40.36%					
Query Match: 25.09%					

US-09-972-758A-2 (1-359) x ABL06282 (1-3714)

```

OY 79 CysProGluSerSerCysLeuArgGluGlyLysGlyGlnAsnGlyAsp----- 95
DB 2507 TGCAGTGGGCGAGAGTTGATTTTGGAGATTTTGGCGGTTCAAAFTTGAATTTGGCTCTC 2448
OY 96 -----AspSerSerAlaGlyGlyAsp 102
DB 2447 TTGCTCTTTCAGAAAGTGGCTCCCAACAGACCTTTGGATACGGCGGAGGAGGGGC 2388
OY 103 PheProProAlaGluValGluProThrProGluAlaGluLeuAlaGlnProCys 122

```

Db 2388 ----- 2388
QY 123 Hisaspsergluaaserlyslenglyalaproalialaglygluglugluproly 142
Db 2387 -----GCCAGTGGCGGAGGCGGTCTCTAGAGGCGGCTCGGAATCCG 2340
QY 143 Glnglnclnaraglulenglylsylslyshisargrargproserlyslslyarg 162
Db 2339 AAGGAGAACATCGCGCGGAAAAATCC---AAGATGACGCCCAAGAACACCAAGAAC 2283
QY 163 Histrlpsprotyrtyrlyslsleuthrtrpdlu-----Gluylslyslslyphe 178
Db 2282 CATATAT---CCGCGATGGAAATCGATATGTCAACGGCGTGCAGAGCAACCGCTGAGGCG 2226
QY 179 Asppluylsglnserleuaragalaserarglleargrlaglmetprialalysglyln 198
Db 2225 AATCAACGACAGATACAGAACCAACCTGCTTCCTCCGCCCTCTTG----- 2178
QY 199 Provalalaprotyrasnthrthrlnpheelumetaspshisaspjlnuglupro 218
Db 2177 ---CTGCTTCGCTACACAGCAACCGCTCTCTATGAGAGCAACATGTCGAG----- 2127
QY 219 AspleulsthrnglyleutyrrserlysaralalalalysSeraspsthrSerasp 238
Db 2126 -----CTGCACAAAGATGACTCCGACGAC 2103
QY 239 Aspasphemetglugluglyglugluaspjlyglyserraspjlymetglylyasp 258
Db 2102 AACGCTTTT-----GGCTCCCAAGCCGAA-----GAT 2076
QY 259 Glyserglupheeluaglnaraspshesercluthrtyrgrluarqyrrhisthrgluser 278
Db 2075 CAGGTGCTCTCTCTCCAGAGAGTCTCCGATGTACAGAGCGGCGCTCGAAGCT 2016
QY 279 Leuclnasmetserlyslnglulengllysluylrleuglulenglulyscysleu 298
Db 2015 TTGGAGACGATGACCAAGAGAGAGCTCATCCAGAGTGCACAAATTCGAATGCATAT 1956
QY 299 Serargmetgluaspgluasnasnargleuargrleugluserlysatrgleuglylyasp 318
Db 1955 TCGAAGGCCAGAACATATCCCAAGAGTTT-----GGAGCCAATATACGAGCGCAG 1905
QY 319 Aspalalargvalaragluenglulenglulenglulenglulenglulenglulenglul 333
Db 1904 GATGATTAAGATGCGCAACATATACGTGACAGAAATCATGATATGCCA 1860

RESULT 12
AAC95217
ID AAC95217 standard; cDNA; 549 BP.
XX
AC AAC95217:
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:1712.
XX
KW Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
KW vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
XX
OS Ctenocephalides felis.
XX
PN W0200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09437.
XX
PR 09-APR-1999; 99US-0128704.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Gaines PJ, Slinchcomb DT, Wisniewski N;

DR WPI: 2000-656323/63.
XX
PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations -
XX
PS Claim 26: Page 818; 964pp; English.
XX
CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
CC anti-HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HNC cDNA of the invention.
XX
SQ Sequence 549 BP; 207 A; 100 C; 113 G; 126 T; 3 other:
SQ
Alignment Scores:
Pred. No.: 0.00489 length: 549
Score: 184.00 Matches: 66
Percent Similarity: 42.29% Conservative: 30
Best Local Similarity: 29.07% Mismatches: 53
Query Match: 9.63% Indels: 78
DB: 21 Gaps: 12
US-09-972-758A-2 (1-359) x AAC95217 (1-549)
QY 34 Proglyalaglulgluargvalprogluagluaspserargtrpnlserargalaph----- 52
Db 33 CCACGCGGTGAGAGAAATAGTG-----AGTAATATCATCATCAAGTGTTTT 80
QY 53 ---Progluengllylaryargprogllyprogluengllylserleugluserglupro 71
Db 81 TTAACTACTGTGAACATCAAAATGACGTGTAAATAATTCGA---AATATGTAATCGGAAT 137
QY 72 Pro-----Proleuglnthrlnalal 78
Db 138 CCAAAATATATACACCGCAGTACAGCGTCCAAATGTCTGCCCGCTTCAACACAGCGCT 197
QY 79 CysprogluserserlysluarggluglylulysglylnasnclysaspserSer 98
Db 198 -----ACTGAGCGGCTCTATCATCGAGAAAGAAAGAAATGATAGTGAT----- 239
QY 99 Alagllylaspheproproproalaglvalgluprothrproglualaglulenglul 118
Db 239 ----- 239
QY 119 AlagluprocysHisaspsergluaaserlyslenglyalaproalialaglyglu 138
Db 240 -----ACAAAGAGCAACCAAGTGGT-----GCAACAAATGGGAG 278
QY 139 Glugluprpgllynglnclnaraglnaraglulengllyslslyslslysarqarqproser 158

QY 223 GlyLeuTyrSerIysArgAlaAlaIaIysSerAsp-----AspThr 236
Db 773 -----ATGGATTTCATCGCGCAAAAGGTGAGTGCAGAGATGCGACCA 826
QY 237 SerAspAspPheMetGluGlu 245
Db 827 GACCCAGACCCAGACACTGCGAGATCGGA 853
RESULT 14
AAK51987
ID AAK51987 standard; cDNA; 7516 BP.
XX AAK51987;
AC
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 532.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001MO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR P-PSDB; AAM78854.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX
PS Claim 1: Page 1918-1926; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM7833-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 7516 BP; 1909 A; 2056 C; 2326 G; 1225 T; 0 other:
Alignment Scores: 1.49 Length: 7516
Pred. No.:

Score: 162.00 Matches: 99
Percent Similarity: 35.73% Conservative: 65
Best local Similarity: 21.57% Mismatches: 183
Query Match: 8.48% Indels: 113
DB: 22 Gaps: 15
US-09-972-758a-2 (1-359) x AAK51987 (1-7516)
QY 2 AlagluProPheLeuSerGluTyrGln-----HisGlnProGln--- 14
Db 2202 GCTGATGCTTACCGCTGAGAGACAGAACCCCACTTTGTCCGCTGCATCTATCCCAACCA 2261
QY 15 -----ThrSerAsnCysThrGlyAlaAlaIaIaGlnGlu 26
Db 2262 CGAAGAAAGGCGCGCAAGCTGACCCGCAFTCTGCTGCTGACACAGCTGGCTTCCAGCA 2321
QY 27 -----GluLeuAsnProGluArgPro-----ProGly 35
Db 2322 TGTTCGAGGGCATCGTATCTGCGCGAGGCTTCCCAACAGAGTGGCTTCCAGCA 2381
QY 36 AlagluGluArgValProGluIuAspSerArgTrrpGlnSerArgAlaPhe-----Pro 53
Db 2382 GTTTCGCGAGAGATATGAGATCTGACTCCAACTCCATTTCCAGAGGTTTCATGGACGG 2441
QY 54 GlnLeuGlyGlyArgProGlyProGluGlySerLeuGlnSerGlnProProPro 73
Db 2442 GAAGCAGCGCTGCTGCTCATGATMAAGCCCTGAGAGCTGCAGCAGCATCTGATCCGCA 2501
QY 74 LeuGlnThrGlnAlaCysProGluSerSerCys----- 84
Db 2502 TGGCCAGAGCAAAAGCTTCTTCCCTGCGGTGCTGGCCACCCTGAGAGAGCGCAGA 2561
QY 85 -----LeuArgGluGlyGlyGln 92
Db 2562 CTTGAAGATACCGCAGCTCATAGAGTTTCCAGGCTGCTGCAGGGCTACTTGGCCAG 2621
QY 93 AsnGlyAspAspSerSerAlaGlyIaIysPhe-----ProProAlaGluVal 109
Db 2622 GAAGCATTTGCCAAGGGGAGCAGCAGCTTACCGCATGAGAGTCTCTCCAGGGAACCT 2681
QY 110 GlnProThrProGluAlaGluLeuAla-----GlnProCysHisAspSerGluAla 127
Db 2682 CGCTGCCCTTACCTAACCTGCGGAACTGCGAGCTGCGCGCTCTTCAACCAAGCTCAAGCC 2741
QY 128 SerIysLeuGlyAlaProAlaIaIaGlyGluGluGluTrp-----GlyGln 143
Db 2742 GCTGCTGCAGAGTACCGGAGCGAGAGAGATGATGATGCCCAAGAGAGAGAGCTGGTGA 2801
QY 144 GlnGlnArgGln-----LeuGlyIysIysHisArgArgArgProSerIysIys 160
Db 2802 GGTACAGAGAAAGACACCTGCTGCGGAGAAACAGCTCACGGAATGGAGACUCCTGCAGTC 2861
QY 161 LysArgHisTrp-LysProTyrTyrIysLeuThrTrpGluGluIuLysIysPheAspG 180
Db 2862 TCAGCTCATGCGCAGAG-----AAATTGACGTCGACAGAGACACTCCAGCAGAAAC 2912
QY 180 uLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaIysGlyGlnProva 200
Db 2913 CGAGCTGTGTGCGGAGCTGAGAGCTCCGGGCCGCTGACCCCAAGAGATGAGAATT 2972
QY 200 lAlaPro----- 202
Db 2973 AGAAGAGATCTGCCATGACCTAGAGGCCAGGCTGGAGAGAGAGAGAGCGTCCAGCA 3032
QY 203 -----TyrAsnThrTrpGlnPheLeuMetAspAspHisAs 214
Db 3033 CCTGACGCGCGAGAGAGAGAGATGACAGAGAAATCCAGAGCTTGAGAGCACTGCA 3092
QY 214 pGlnGluGluProAspLeuIysThrGlyLeuTyrSerIysArgAlaIaIaIysSerAs 234
Db 3093 GAGAGAGAG-----AGCGCCCGCGAGAAAGCT-GCAGCTGAGAGAAAGGTGA 3136
QY 234 pAspThrSerAspAspPheMetGluGluGlyGlyGluIuAspGlyIysSerAspG 254

Db 1030 GTGCAACCTGGAGAACGAGAAAGACTTTGACCAGCTCTCGCGGAGAGAACCAT 1089
OY 118 LeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaPro----- 133
Db 1090 CTCTGCGCAAGTATGACAGAGAGCGGACCGGGCTGAGCGGAGCGCCGAGGAAGAGAC 1149
OY 134 -----AlaAlaGlyGlyGlu 139
Db 1150 CAAGGCTCTGCTGCGCGCGCGCGCTGGAGAAAGCCATGGAGCGAAGGCGGACTGGAG 1209
OY 140 GluTrpGlyGlnGlnArg-----GlnLeu 148
Db 1210 CGCCTCAACCAAGCATTCGCCGACGAGATGACACTTATGACCTCCAAAGATGATGTG 1269
OY 149 GlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyr 168
Db 1270 GGCAAGAGTGTCCAGACAGCTGGAGAGTCCAAAGCGGCCCTAGACACAGCAGGTGAGAG 1329
OY 169 LysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArg 188
Db 1330 ATGAAGACGCGAGCTGGAGAGCTGAG--GACGAGCTGACGGCCAGCAGATGCCAAG 1386
OY 189 IleArgAlaGluMetPheAlaLysGlyLysProValAlaProTyrAsnThrThrGlnPhe 208
Db 1387 CTGCGGTTGGAGGTCACCTGACGCGCATGAAAGCC----- 1422
OY 209 LeuMetAspAspHisAspGluGluGluProAspLeu-LysThrGlyLeuTyrSerLysAr 228
Db 1423 -----CACTTGGAGCGGAGCCTGCGAGGCGCGGAGCAGACAGCGAG 1464
OY 228 GAlaAlaAlaLysSer-AspAspThrSerAspAspPheMetGluGluGlyGlyGlu 248
Db 1465 GAGAAAGAAAGCAGCTGTGTCACAGAGTCCGAGATGAGGAGCAGCTGGAGAGCAG 1524
OY 248 LuAspGlyGlySerAspGlyMetGly-----GlyAspGlySerGlu-Phe 262
Db 1525 AGGAGGACGCGCTGTGAGCGAGTGGCGCGCGGAGAAAGCTGAGATGACCTGAAGGAC 1584
OY 263 LeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMet 282
Db 1585 CTGGAGGCGGCACATC-----GACTCGGCCAACAGAACCGGAGAGACCATCAACAGCTG 1641
OY 283 -----Ser 283
Db 1642 CGGAAGCTGCAGGCCAGATGACAGACTGCATGCGGAGCTGATGACACCCGCGCTCT 1701
OY 284 LysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAsp 303
Db 1702 CGTGAGGAGATCTGGGCCAGGCCAAAGAGACAGAAAGACTTGAAGAGCATGAGAGCC 1761
OY 304 GluAsnAsnArgLeuArgLeuGluSer----- 312
Db 1762 GAGATGATCCAGTTGCAGAGAGAACTGGCAGCCGCGAGCTGCCAAGCGCCAGGCCAGC 1821
OY 312 ----- 312
Db 1822 AGGAGCGGATGAGCTGGCTGACGAGATCGCCACACAGAGCGGCAAGAGCCCTGGCGT 1881
OY 313 ---LysArgLeuGluGlyAspAspAlaArgValArgGluLeuGluLeu----- 329
Db 1882 TAGAGGAGAGCGGCGTCTG-GAGGCCCGCATCGCCAGCTGAGAGAGAGCTGGAGAG 1940
OY 330 -----AspArgLeuArgAlaGluAsnLeuGlnLeu--- 339
Db 1941 GAGCAGGGCAACAGGAGCTATACAGACCGGCTGAAGAAGGCCAACTGCAGATCCAG 2000
OY 340 -----LeuThrGluAsnGluLeuHisArg 347
Db 2001 CAGATCAACACCGACTGAACCTGGAGCGGAGCGCCGAGAAAGAGAAATGCTGG 2060
OY 348 GlnGln 349
Db 2061 CAGCAG 2066

Search completed: June 17, 2003, 10:54:16
Job time : 292 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 17, 2003, 09:52:47 ; Search time 60 Seconds
(without alignments)
1834.949 Million cell updates/sec

Title: US-09-972-758A-2

Perfect score: 1910
Sequence: 1 MAEPFLSEYQHOPQTSNCTG.....LTENELHROORAPLSKFGD 359

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09972758/runatc.17062003.095238.24394/app_query.fasta.1.519
-DB=Issued Patents.NA -OFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09972758.@CGN.1.1.40.@runatc.17062003.095238.24394 -MCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONCLG
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents.NA:*
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
4: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
5: /cgn2.6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	158.5	8.3	9551	US-08-056-200-93	Sequence 93, Appl
2	158.5	8.3	9551	US-08-800-644-93	Sequence 93, Appl
3	155	8.1	2277	US-08-676-967-5	Sequence 5, Appl
4	155	8.1	2277	US-08-676-967-5	Sequence 5, Appl
5	155	8.1	2277	US-09-098-487-5	Sequence 5, Appl
6	141	7.4	2304	US-08-464-266-1	Sequence 1, Appl
7	141	7.4	2304	US-08-464-272-1	Sequence 1, Appl
8	141	7.4	2304	US-08-464-514-1	Sequence 1, Appl
9	141	7.4	2304	US-08-486-403-1	Sequence 1, Appl
10	137	7.2	1617	US-09-265-013-2	Sequence 2, Appl
11	137	7.2	7308	US-09-011-745-3	Sequence 3, Appl
12	137	7.2	7308	US-09-011-745-4	Sequence 4, Appl

13	137	7.2	7616	US-09-011-745-2	Sequence 2, Appl
14	137	7.2	8202	US-08-258-420-13	Sequence 13, Appl
15	137	7.2	8332	US-08-850-961-1	Sequence 1, Appl
16	137	7.2	8332	US-09-479-776-1	Sequence 1, Appl
17	137	7.2	8332	US-09-309-572-11	Sequence 11, Appl
18	137	7.2	8332	US-09-315-127-1	Sequence 1, Appl
19	137	7.2	8332	US-09-265-013-1	Sequence 1, Appl
20	134	7.0	3930	US-09-162-373-2	Sequence 2, Appl
21	134	7.0	3930	US-09-467-946-2	Sequence 2, Appl
22	133.5	7.0	6755	US-08-931-999-4	Sequence 4, Appl
23	131.5	6.9	5661	US-08-938-105-2	Sequence 2, Appl
24	131	6.9	3023	US-09-593-889-10	Sequence 10, Appl
25	128.5	6.7	2581	US-09-370-838-66	Sequence 66, Appl
26	126	6.6	7791	US-08-949-386-23	Sequence 23, Appl
27	126	6.6	7791	US-08-450-562-22	Sequence 22, Appl
28	126	6.6	7791	US-08-450-562-22	Sequence 22, Appl
29	126	6.6	7791	US-08-984-709A-23	Sequence 23, Appl
30	126	6.6	7808	US-08-450-272-23	Sequence 23, Appl
31	126	6.6	7808	US-08-149-097D-22	Sequence 22, Appl
32	126	6.6	7808	US-08-949-386-22	Sequence 22, Appl
33	126	6.6	7808	US-08-450-562-22	Sequence 22, Appl
34	126	6.6	7808	US-08-984-709A-22	Sequence 22, Appl
35	125.5	6.6	2384	US-07-814-964-10	Sequence 10, Appl
36	125.5	6.6	2384	US-08-258-442-10	Sequence 10, Appl
37	125.5	6.6	2384	US-08-328-809-5	Sequence 5, Appl
38	125.5	6.6	2384	PCT-US92-11107-10	Sequence 10, Appl
39	125	6.5	2672	US-09-214-564A-5	Sequence 5, Appl
40	125	6.5	2815	US-09-214-564A-1	Sequence 1, Appl
41	124.5	6.5	1276	US-09-177-325-2	Sequence 2, Appl
42	124.5	6.5	1276	US-09-411-812A-2	Sequence 2, Appl
43	124.5	6.5	1276	US-09-590-113-2	Sequence 2, Appl
44	124	6.5	4722	US-08-979-608A-14	Sequence 14, Appl
45	122	6.4	3201	US-08-556-419-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-056-200-93
Sequence 93, Application US/08056200
Patent No. 5616500
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/056,200
FILING DATE: 30-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502

```

: INFORMATION FOR SEQ ID NO: 93:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 9551 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: cDNA
:   HYPOHETICAL: NO
:   ANTI-SENSE: NO
:   FEATURE:
:     NAME/KEY: CDS
:     LOCATION: 1507..1644
:     FEATURE:
:     NAME/KEY: Intron
:     LOCATION: 1645..2511
:     FEATURE:
:     NAME/KEY: CDS
:     LOCATION: 2512..8070
:
: US-08-056-200-93
:
: Alignment Scores:
:   Pred. No.: 0.00154
:   Score: 158.50
:   Percent Similarity: 39.89%
:   Best Local Similarity: 25.56%
:   Query Match: 8.30%
:     DB: 1
:           Gaps: 16
:
: US-09-972-758a-2 (1-359) x US-08-056-200-93 (1-9551)
:
: QY      31  GUUArgProProGluAlaGluGluArg-----ValProGluGlu 43
:         |||||  |||||  |||||  |||||  |||||  |||||  |||||
: Db      4810  GAGAGGAGCGTGAAGCGCGGAGCGCGACAGCGGGAACACAGGTTTCTCCCGGAGAGG 4869
:
: QY      44  AspSerArgTrpPheIleSerArgAlaPheProGlnLeuGlyIleArgProGlyProGluGly 63
:         :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
: Db      4870  GAGGAGAAAGGAGGAGCGC-----GGCGCGCAGCGAGCGGAGAGG 4908
:
: QY      64  GluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysProGlnSerSer 83
:         |||  |||||  |||  |||  |||  |||  |||  |||
: Db      4909  GAAAGAAGAGCTGACGTTCCCTGAGAGAAAGAGAGACAGCTCCAGCGCGGAGCGCTGCCCA 4968
:
: QY      84  CysLeuArgGluGluGlyIleGlyGly---GlnAsnGlyAspSerSerAlaGlyIleAsp 102
:         |||:|:|:|  |||  |||  |||  |||  |||  |||  |||
: Db      4969  CACCTCCAGAGAGAGAGGAGCGCGCTCCACGA-GCATCAGAGAGAGAGCGCGAGAGAGCA 5027
:
: QY      103  PhePro---ProProAlaGluVal-----GluProThrProGlu 114
:         |||  |||  |||||  |||||  |||||  |||||  |||||
: Db      5028  GCGCGCGGACCAAAATGGAGGTGGCACTAGAAAGAAAGAAAGAGACCGCCGACAC 5087
:
: QY      115  AlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerIleArgIleAlaProAla 134
:         |||  |||  |||  |||  |||  |||  |||  |||
: Db      5088  GCGTGTACGCCACACCGCTACAGAGCAGCTGAGAGAGAAC-----GCAGCGCT 5141
:
: QY      135  AlaGlyGly-GluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyIleLysLysLysHisAr 154
:         |||||  |||  |||||  |||  |||||  |||||  |||||
: Db      5142  GCAGAGAGAGAGAGAGAGCTTACAGAGAGAGAGAGCGCGAG-----AA 5183
:
: QY      154  GATGATGProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGlu 174
:         :|||:|:|:|  :|||:|:|:|  :|||:|:|:|  :|||:|:|:|
: Db      5184  GAGAAAGCGCGCAAGAACAGAGACATACCCGAGAGAGAGAGAGAGAGCTGCAGAGAGAGGA 5243
:
: QY      174  uLysLysLysPheAspGlnLysGlnSerLeuArgAlaSerArgIleArgIleAlaLumetPh 194
:         |||||  |||||  |||||  |||||  |||||  |||||  |||||
: Db      5244  AGAGCAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5303
:
: QY      194  eAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetLysAspHisAs 214
:         :|||  :|||  :|||  :|||  :|||  :|||  :|||
: Db      5304  TCGAAGCATTAAGAG-----CTGACGACGAGAGAGAGA 5336
:
: QY      214  pGln-----GluGlnProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAl 231
:         :|||  :|||  :|||  :|||  :|||  :|||  :|||
: Db      5337  GCAGCTGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5381
```

```

: QY      231  aLysSerAspAspThrSerAspAspPheMetGluGluGlyGluGluAspGly 251
:         :|||  :|||  :|||  :|||  :|||  :|||  :|||
: Db      5382  GGAGAAAAAATACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5421
:
: QY      251  ySerAspGlyMetGlyGlyAspGlySerGlnPheLeuGlnArgAspPheSerGluThrTy 271
:         |||  |||  |||  |||  |||  |||  |||
: Db      5422  -----GAGAGCAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5456
:
: QY      271  rGluArg-----TyrHisThrGlnSerLeuGlnAsnMetSerLysGlu 285
:         |||  |||  |||  |||  |||  |||  |||
: Db      5457  AAGCGCCAGAGCTGGAGAGAGCGCATCCGCAAAAAGAGAGAGCTGCACAGAGAGAGAGA 5516
:
: QY      285  nGluLeuIleLysGluTyrLeuGlnLeuGlnLysCysLeuSerArgMetGluAsp----- 303
:         :|||:|:|:|  :|||:|:|:|  :|||:|:|:|  :|||:|:|:|
: Db      5517  GCAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5576
:
: QY      304  -GluAsnAsnArgLeuArgLeuGlnLysSerLysArgLeuGlyIleAspAspAlaArgValAr 323
:         |||  |||  |||  |||  |||  |||  |||
: Db      5577  GGAGAGAAAGAGAGAGCTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5636
:
: QY      323  g-----GluLeuGlnLeuGlnLeu-----AspArgLeuArgAlaGluAsnLe 337
:         :|||  :|||  :|||  :|||  :|||  :|||  :|||
: Db      5637  AAGCGCCAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5696
:
: QY      337  uGlnLeuLeuThrGluAsn---GluLeuHisArgGlnGlnGluArg 351
:         |||||  |||  |||  |||  |||  |||  |||
: Db      5697  GCAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5742
:
: RESULT 2
: US-08-644-93
: Sequence 93, Application US/08800644
: Patent No. 5958752
:
: GENERAL INFORMATION:
:   APPLICANT: Steinert, Peter M.
:   APPLICANT: Lee, Seung-Chul
:   APPLICANT: Kim, In-Gyu
:   APPLICANT: Chung, Soo-Il
:   APPLICANT: Park, Sang-Chul
:   TITLE OF INVENTION: Trichomyalin and Transglutaminase-3 and
:   TITLE OF INVENTION: Methods of Using Same
:   NUMBER OF SEQUENCES: 117
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Knobbe, Martens, Olson & Bear
:     STREET: 620 Newport Center Drive, Sixteenth Floor
:     CITY: Newport Beach
:     STATE: CA
:     COUNTRY: U.S.A.
:     ZIP: 92660
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: Floppy disk
:     COMPUTER: IBM PC compatible
:     OPERATING SYSTEM: PC-DOS/MS-DOS
:     SOFTWARE: Patentin Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/800,644
:     FILING DATE: 14-FEB-1997
:     CLASSIFICATION: 424
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US 08/056,200
:       FILING DATE: 30-APR-1993
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Fedrick, Michael F.
:     REGISTRATION NUMBER: 36,799
:     REFERENCE/DOCKET NUMBER: NIH054,001A
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: (714) 760-0404
:       TELEFAX: (714) 760-9502
:   INFORMATION FOR SEQ ID NO: 93:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 9551 base pairs
:       TYPE: nucleic acid
:       STRANDEDNESS: single
```



```

OY      6 LeuSerGIuTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAlaValaGln 25
DB      144 CTAGGTGACCTTCACAT-----GCTGGAGAGACGTGCACCGCGC 182
OY      26 GlnGlu-----LeuAsnProGlnArgProProGlyAlaGln 38
DB      183 CCTGAGGAGATCACACCTTCGAGGCTGCACAGATGACCTGACCGCCCAAGAA 242
OY      39 ArgValProGlnGluAspSerArgTrpInSerArg-----AlaPheProGlnLeu 56
DB      243 GCTGGCCACACAGACCAAGAGAAAGGAGAGACAGACAGACAGCTGCCCAAGAA 302
OY      57 GlnArgProGlyProGlnGlyGlySerLeuGlnSerGlnPro----- 71
DB      303 GCCCAA---GCCCAAGAGAGCCAGGTGCGCAGACAAAGAGCCCGCCATCATCCGCA 359
OY      72 ProProLeuGlnThrGlnAlaCysProGlnSerCysLeuArgGlnGlyGly 91
DB      360 CCTGAGCTTCAAGTCCAGCCGAGACGACCTGAGACCGCTGTCCCGCAGTTCGGGC 416
OY      92 GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaGlnValaGlnPro 111
DB      417 -----CGTCTGGAGGTGAACATCCCGCAAGCCGACGCCAA--- 455
OY      112 ThrProGlnAlaGlnLeuLeuAlaGlnProCysHisAspSerGlnAlaSerLysLeu 131
DB      456 -----GATGCGCGCTTCGGCTTCGTGCAGTTCAGAAACCTGCTGAGCGCGCAAGC 509
OY      132 AlaProAlaAlaGlyGlyGlnGlu----- 140
DB      510 -----CCTGAAGGACATGAACATGAAGAGATCAAGGGCCGACCGTGGCTGACTG 563
OY      141 -----TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysHisArgArg-ArgPr 157
DB      564 GGCGGTGGCCAAAGGACAAAGTACAAAGACACCCAGAGCTGAGCGCATCGCGAGAGAA 623
OY      157 oSerLysLysLysArgHisTrpLysProTyrTrpLysLeuThrTrpGlnGlnLysLys 177
DB      624 GAGCCACAGAGAGACACCAAGAGAGAGCGTCAAGAAAGCCCGCAGAGAGAGAT 683
OY      177 sPheAspGlnLysGlnSerLeuArgAlaSerArgIleArgAlaGlnMetPheAlaLysG 197
DB      684 GGAGAGAGAGAGAAACGAC----- 702
OY      197 GlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGln 217
DB      703 -GAGCAGCAGCAGCAGCAGCAGCAGAGAGAGCGGCTGTTCAGCAGCAGCAGAGAGA 761
OY      217 uProAspLeuLysThrGlyLeu-----TyrSerLysArgAlaAla---- 230
DB      762 GGAGACATCGAGCAAGGTGACCAAGCCCGTCAGATCCAGAAAGCCCGCGTGAAGC 821
OY      231 -----AlaLysSerAspAspThrSerAspAspAspPhe---MetGlnGlnGly 246
DB      822 CCCCCCCCCCAAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 881
OY      246 YGlnGluAspGlySerAspGlyMetGlyGlyAspGlySerGlnPheLeuGlnArgAs 266
DB      882 CAGCATGAC-----GAGCGCAGAGAGCTGGCCACAGAGCA 917
OY      266 pPheSerGlnThrTrpGlnArgTyrHisThrGlnSerLeuGlnAsnMetSerLysGln 285
DB      918 CACACGACACCGAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 977
OY      285 ----- 285
DB      978 GCTGCCAGCAGCGTGAACAGAGCGCAGTTCATCCGCAACCTGACCTTCGACAG 1037
OY      286 -----GlnLeuIleLysGlnTyrLeuGln----- 294
DB      1038 CGAGAGAGAGAGCTGGCGGAGCTGCTGCAGACAGTTCGGCGAGCTGAAGTACGTGCGCAT 1097

```

```

OY      294 ----- 294
DB      1098 CGTCTGACCCCGACACCGAGCAGACAGAGGCTGGCGCTTCGCCCATCATGACCCA 1157
OY      295 -----GlnLysCysLeu-----SerArgMetGlnAspGlnAsnArgLeuArg 309
DB      1158 GGAGCGCGCCGACAGAGTCTGCTGGCGCCGAGACCGCCGAGAGAGCGCGCGCTGAA 1217
OY      309 GlnGlnSerLysArgLeuGlnGlyAspAspAlaArgValaArgGlnLeuGlnGln 329
DB      1218 GCTGACCGCGCCAGCTG-----AAGGTGACCTGGCGCT 1253
OY      329 uAspArgLeuArgAlaGlnAsnLeuGlnLeuThrGlnAsnGlnLeuHisArgGlnG 349
DB      1254 GACCGCGCAGCAGCGCGCCGACAGCTGCAGACCAAGGT-CAAGAAAGCCGACCGCAGCC 1312
OY      349 nGlnArgAlaPro 353
DB      1313 GCAACCTGTACT 1325

RESULT 4
US-08-676-974-5
: Sequence 5, Application US/08676974
: Patent No. 5770422
: GENERAL INFORMATION:
: APPLICANT: COLLINS, KATHLEEN
: TITLE OF INVENTION: Human Telomerase
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Science & Technology Law Group
: STREET: 268 Bush Street, Suite 3200
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/676,974
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman Ph.D., Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UCB96-055
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415)343-4341
: TELEFAX: (415)343-4342
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2277 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-676-974-5

Alignment Scores:
Pred. NO.: 0.000421 Length: 2277
Score: 155.00 Matches: 96
Percent Similarity: 33.93% Conservative: 55
Best Local Similarity: 21.57% Mismatches: 147
Query Match: 8.12% Indels: 148
DB: 1 Gaps: 18

US-09-972-758A-2 (1-359) x US-08-676-974-5 (1-2277)
OY      6 LeuSerGIuTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAlaValaGln 25
DB      144 CTAGGTGACCTTCAGAT-----CCTGGAGAGACGTGCAGCGCGC 182

```


QY 13 ProGlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGluLeuAsnProGluArg 32
111 : : : : :
388 CCCAATTGGCTGAGAGCTCTGCCCGCTGACATCGACAGAGATGCCCATCAACAT 447
QY 33 ProProGlyAlaGluArg-----ValProGluLysSerArgTyrPglSer 49
111 : : : : :
448 CCGGTGAGCGGACGACACCTCTGCTTATTTG--CGGGATTCGGCCAGTGGCAACA 506
QY 50 ArgAlaPheProGlnLeuGlyGlyArgProGly-----ProGluGlyGly 65
111 : : : : :
507 CTACGGCGGTACAGCTGTGAGGGCTGCAAGGCTTTTAAACGCACACTGCCAAGA 566
QY 66 Ser-----LeuGluSerGlnProProProLeuGlnThrGlnAlaCysProGluSer 82
111 : : : : :
567 TCTCATACATACGCTTCCAGGGAGAACCCCACTCATACACAGCGCGCAGAGAA-- 623
QY 83 SerCysLeuArgGluGlyGlyGlyGlyGlnAsnGlyAspSerSerAlaGlyLysP 102
623 -----
QY 103 PheProProAlaGluValGluProThrProGluAlaGluLeuAlaGlnProCys 122
624 -----CCGCTGCCAGTACTGCGCTACCAAGAGTGCCTTAACCTGCGG----- 665
QY 123 HisAspSerGluAlaSerLysLeuGlyAla-----ProAlaAlaGlyGlyGlu 138
111 : : : : :
666 CATGAACCGCGAAGCGCTCCAGAGAGAGCGCTCAACCGCGCGCGCAATGCGGCGG-- 722
QY 139 GluGluTyrPglGlnGlnGlnArg----- 146
723 -----TAGGCTAGCGCCAGCGGAGCGGAGCGGAGTACAGGCTTCGCTAGCGGATC 776
QY 147 GlnLeuGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 166
111 : : : : :
777 CAGCTTCACAGCGCGAGAGAGAGAGAGCGG-----CCTTCTGCGGGAAT 821
QY 167 TyrTyrLysLeuThrTyrPglGluLysLysLysPhe-AspGluLysGlnSerLeuArgAl 186
822 -----GGCAGCGCGCAACGCTTGTGATGACTCATGACCAATACCGT 863
QY 186 AsnArg-----IleArgAlaGluMetPheAlaLysGlyGln-- 198
864 GTCCAGGAGATTTCATCGATCGAGCGCATATGAGAGCGGAGCGGAGCCCAATG 923
QY 199 -----ProValAlaProTyrAsnThrThrGlnPheLeu 210
924 CGGCGATCGTCACGTACGTTCTCGCGGCTTGCTTATTCACAGTCCAG----- 975
QY 210 LaspAspHisAspGlnGluProAspLeuLysThrGlyLeu----- 224
976 -----CCGGACTACAGGGTGGCGTGTGCGGCTGTGCCAAGT 1013
QY 225 -----TyrSerLysArgAlaAlaLysSerAs 234
1014 GGTCAACAACAGCTCTTCAGATGTGTGAATAGCGCGGATGATGCCGCACTTTGCCA 1073
QY 234 PaspThrSerAspAspAspPhe----- 241
1074 GGTGGCGCTGGAGAGACAGTGATTCGTGAAAGCCGCTTGATCGACTGCTCATTCG 1133
QY 242 -----MetGluGluGlyGlyGluLysPglLysPgl 252
1134 GAACGTGGCTGTGTCAGCATCTTTCGTTGCTGTGATGAGCGGCT--GCCGCGCGGCGG 1190
QY 252 RasPglMetGlyGlyAspGlySer----- 260
1191 CGGTGAGCTAGGCGACAGCTCTTTTGAGCAGATCACCGGCGCTTACGCCCAACA 1250
QY 261 -GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrLysLeu-- 279
1251 GGTGTCTCTCACACAGAGCTTCTCG-----TACCATGGCAACAGTGGCAT 1295
QY 279 ----- 279

Db 1296 CAAAGCCGCTGTGCACCATCTTCGACCCATATTCGAGAGTAGTGTAAGATGAA 1355
QY 280 -----GlnAsnMetSerLysGlnGluLeu----- 287
1356 GCGGCTGAATTCGTGACGAGCGAGCGTCTGCTTCGTAAGCCATCATGTAACAACC 1415
QY 288 -----IleLysGlnTyrLeuGlnGluLysCysLysSerArgMetGlu-- 302
1416 GGACATACGGCGGATCACAGCGCGGAGATGAGATGAGATGCCCCAGAGAGCTGACGC 1475
QY 303 -----AspGluAsnArgLeuArgLeuGluSerLysArgLeuGlyLysAspAl 320
1476 TTGGCTGACGAGACACTGCGCGCTGGAACATCCG-----GCCAGCATGG 1520
QY 320 ArgValArgGlnLeuGlnGluLeuAspArg-LeuAlaGlnAlaGlnLeuGlnLeu 340
1521 ACGCTTGGCCCACTGCTGCTGCGTGTGCGCCCTTGGCATGATCAGCTGAAGTGC 1580
QY 340 eutHrgLuanGlnLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 356
Db 1581 AGCATCACTGTTCTCTTCGCAATTTACAGAGCGCGCGCTGAGAGAG 1630
RESULT 7
US-08-464-272-1
Sequence 1, Application US/08464272
Patent No. 5686691
GENERAL INFORMATION:
APPLICANT: ORO, Ph.D., ANTHONY E.
APPLICANT: EVANS, Ph.D., RONALD M.
TITLE OF INVENTION: INSECT RETINOID-LIKE RECEPTOR
TITLE OF INVENTION: COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: United States
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,272
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/013,975
FILING DATE: 04-FEB-1993
APPLICATION NUMBER: US 07/497,935
FILING DATE: 22-FEB-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reltier, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9350
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2304 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..1704
US-08-464-272-1

Alignment Scores:

Pred. No.: 0.00669 Length: 2304
 Score: 141.00 Matches: 106
 Percent Similarity: 32.43% Conservative: 49
 Best Local Similarity: 22.18% Mismatches: 127
 Query Match: 7.388 Indels: 197
 Gaps: 22

US-09-972-758a-2 (1-359) x US-08-464-272-1 (1-2304)

```

OY 13 ProGlnThrSerAsnGlyThrGlyAlaAlaAlaValGlnGlnGluLeuAsnProGluArg 32
DB 388 CCAATTCGCTGAGAGCTCTCCGCGCTGCAGCCAGCCAGATATCCCTTAACCAT 447
OY 33 ProProGlyAlaGluGluArg-----ValProGluGluAspSerArgTrpGlnSer 49
DB 448 CCGGTGAGCGGAGAGACACCTCTGCTCTATTG-CGGGATCGGGCGCATGGAGACA 506
OY 50 ArgAlaPheProGlnGlyArgProGly-----ProGluGlyGlu 65
DB 507 CTACGGCGGTACAGCTGTGAGGCTGCAGGGCTTCTTAAACGCACAGTGGCAGAGA 566
OY 66 Ser-----LeuGlnSerGlnProProGluGlnThrGlnAlaCysProGluSer 82
DB 567 TCTACATACCTTGCAGGAGAGAACCCAGCATCATAGACACGCGGAGAGAA-- 623
OY 83 SerCysLeuArgGluGluGlyGluGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102
DB 623 ----- 623
OY 103 PheProProAlaGluValGluProThrProGluAlaGluLeuAlaGlnProCys 122
DB 624 -----CCGCTCCAGCTACTCGCTCCAGCAAGTCAACTCGC----- 665
OY 123 HisAspSerGluAlaSerLysLeuGlyAla-----ProAlaAlaGlyGlyGlu 138
DB 666 CATGAGCGCGAGCGGCTCCAGAGAGAGCTCAACCGCGCCGCAATCGCGGGG-- 722
OY 139 GluGluTrpGlyGlnGlnArg----- 146
DB 723 -----TAGGCTCAGCGCCAGCGAGCGGCGCATAGCGTTCGTAAGCGGATC 776
OY 147 GlnLeuGlyLysLysLysHisArgArgProSerLysLysLysAlaGlnHisTrpLysPro 166
DB 777 CAGTCTCCAGGCGGAGAGAGAGAGGGG-----CGTTCTGGCGGAGAT 821
OY 167 TyrTrpLysLeuThrTrpGluGluLysLysLysPhe-AspGluLysGlnSerLeuArgAl 186
DB 822 -----GGGACAGCGGCAACGCTTGTGATGACTTCAATGACCATAGCGT 863
OY 186 AsnArg-----LeuArgAlaGluMetPheAlaLysGlyGln-- 198
DB 864 GTCCAGGATTTTCGATCGAGCGCATCATAGAGCGAGAGCGGAGAGACCATG 923
OY 199 -----ProValAlaProTyrAsnThrThrGlnPheLeu 210
DB 924 CGGCGATCGTCACTGACGTTCTCGCGGTTGTGCTCTATTCCACAGTCAG----- 975
OY 210 LaspAspHisAspGlnGluGluProAspLeuLysThrGlyLeu----- 224
DB 976 -----CCGACTACAAAGGCGCTGTGCGCGCGCTGTGCCAAGT 1013
OY 225 -----TyrSerLysArgAlaAlaAlaLysSerAs 234
DB 1014 GGTCAACAACAGCTTTCAGATGTCGATACGCGGCGATGATGCGGCACTTGGCCA 1073
OY 234 PAspThrSerAspAspAspPhe----- 241
DB 1074 GGTGCGCTGCAGACACAGCATGATTCGTCAAAAGCGCTTGATGAGCTCATTC 1133
OY 242 -----MetGluGluGlyGlyGlyGluGluLysGlyGlySe 252
DB 1134 GAACGTGCGCTGTCAGACATCTTTTGGCTGTGATGAGCGGCGT--GCCGCGGCGGCG 1190

```

```

OY 252 LaspGlyMetGlyGlyAspGlySer----- 260
DB 1191 CGGTGAGATAGCCACCATGCTCTCTTTGACGACGATCACCAGGCGCTTACGCCACACA 1250
OY 261 -GluPheLeuGlnArgAspPheSerGluThrTrpGluArgTyrHisThrGlnSerLeu-- 279
DB 1251 GCTGTTCCTCAACAGACCTTCTCG-----TACCATGCAACACTGGCAT 1295
OY 279 ----- 279
DB 1296 CAAGCCGCTGTGACCATCTTCGACCCGATATTGTGCGAGCTGAGTGAAGTGA 1355
OY 280 -----GlnAsnMetSerLysGlnGluLeu----- 287
DB 1356 GCGGTGATATCTCAGCCAGCGAGGCTGTCTGCTGAAGCCATCATCTGTAACAACC 1415
OY 288 -----LysGluTrpLysGluGluGluGluLysCysLysSerArgMetGlu-- 302
DB 1416 GCACATACGCGCGATCAAGACCGCGCGAGATGAGATGCGCGAGAGAGTGTACGC 1475
OY 303 -----AspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAla 320
DB 1476 TTGCTTGAGAGAGACTGCTCCGCTGGAACATCCG-----GGCGAGATGG 1520
OY 320 ArgValArgGluLeuGluGluLeuAspArg-LeuArgAlaGluAsnLeuGlnLeu 340
DB 1521 ACGCTTGCGCACTGCTGCTGCTGCTGCGCGCTTGTGATGATGATGATGATGATGATG 1580
OY 340 eutHrgLysGlnGluHisArgGlnGlnGluArgAlaProLeuSerLys 356
DB 1581 AGGATCACTGTTCTCTTCCGATTTACACAGCAGCGCGCTGAGAGAG 1630

RESULT 8
US-08-464-514-1
: Sequence 1, Application US/08464514
: Patent No. 6265173
: GENERAL INFORMATION:
: APPLICANT: EVANS, RONALD M.
: APPLICANT: MCKEOWN, MICHAEL B.
: APPLICANT: ORO, ANTHONY E.
: APPLICANT: SEGRAVES, WILLIAM A.
: APPLICANT: YAO, TSO-PANG
: TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE
: TITLE OF INVENTION: STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PRETTY, SCHROEDER, BRUEGEMANN & CLARK
: STREET: 444 South Flower Street, Suite 2000
: CITY: Los Angeles
: STATE: California
: COUNTRY: United States
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/464,514
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/907,908
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Reiter, Stephen E.
: REGISTRATION NUMBER: 31192
: REFERENCE/DOCKET NUMBER: P41 9321
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 546-4737
: TELEFAX: (619) 546-9392

```

```

: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 2304 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: DNA (genomic)
:   FEATURE:
:     NAME/KEY: CDS
:     LOCATION: 163..1701
:
: US-08-464-514-1
:
: Alignment Scores:
:   Pred. No.: 0.00669   Length: 2304
:   Score: 141.00   Matches: 106
:   Percent Similarity: 32.43%   Conservative: 49
:   Best Local Similarity: 22.18%   Mismatches: 127
:   Query Match: 7.38%   Indels: 197
:   DB: 4   Gaps: 22
:
: US-09-972-758a-2 (1-359) x US-08-464-514-1 (1-2304)
:
: Oy 13 ProGlnHrSerAaNCyThGlyAlaAlaValGlnGluLeuAsnProGluArg 32
:   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 388 CCCAATTCGGCTGAGCGCTCTGCCCGCTGACGTCAGCGAGTATCCCGCTAACCAT 447
:
: Oy 33 ProProGlyAlaGluArg-----ValProGluGluAspSerArgTrpGlnSer 49
:   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 448 CCGCTGACGGCAGACCAAGACCTCTGCTATTTTG-CGGGATCGGGCCATGGCAACA 506
:
: Oy 50 ArgAlaPheProGlnLeuGlyGlyArgProGly-----ProGluGlyGlyGly 65
:   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 507 CTACGGCGTGACAGCTGTGAGGCTGCACAGCGCTTTTAACGCCACAGTGGCAAGA 566
:
: Oy 66 Ser-----LeuGluSerGlnProProProLeuGlnTrpGlnAlaCysProGluSer 82
:   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 567 TCTCATACGCTTGCAGGAGGAGAACCGCACTGCATATAGCAAGCGCGAGAGGA--- 623
:
: Oy 83 SerCysLeuArgGluGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAsp 102
:   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 623 ----- 623
:
: Oy 103 PheProProAlaGluValGluProThrProGluAlaGluLeuAlaGlnProCys 122
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 624 -----CCGCTGCCAGTACTGCCGCTACCAGAAAGCTTAACCTGCG- 665
:
: Oy 123 HisAspSerGluAlaSerLysLeuGlyAla-----ProAlaAlaGlyGlyGly 138
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 666 CATGAACCGCGAAGCGCTCCAGAGAGCCTCAACGCGCGCGCCGCAATGCGCGCG- 722
:
: Oy 139 GluGluTrpGlyGlnGlnArg----- 146
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 723 -----TAGCCTCAGCCGCGAGCGGAGCGGAGCGTCCAGTTCCATGACCAATAGCGT 776
:
: Oy 147 GlnLeuGlyLysLysLysHisArgArgProSerLysLysLysArgHisTrpLysPro 166
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 777 CAGCTCTCAAGCGGAGGAGGAGGAGCGCG-----CGTTCTGCGCGGAGAT 821
:
: Oy 167 TyrTrpLysLeuTrpGluGlnLysLysLysPhe-AspGluLysGlnSerLeuAlaGAl 186
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 822 -----GGGCAACGCGCAAGCGTTTCATACCTTCATGACCAATAGCGT 863
:
: Oy 186 AserArg-----IleArgAlaGluMetPheAlaLysGlyGln-- 198
:   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
: Db 864 GTCACGAGATTTCGATCGAGCCATCATAGAGCCGAGCAGGAGCGGAGACCCAAATG 923
:
: Oy 199 -----ProValAlaProTyrAsnThrThrGlnPheLeu 210
:   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
: Db 924 CGGCGATCGTCACTGACGTCTCGCGGTGTCCTATTCACAGTCCAG----- 975
:
: Oy 210 LAspAspHisAspGlnGluProAspLeuLysThrGlyLeu----- 224
:   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
: Db 976 -----CCGACTACAAAGGTGCCGTGTCGCGCCCTGTGCCAAGT 1013
```

```

: Oy 225 -----TyrSerLysArgAlaAlaAlaLysSerAs 234
:   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 1014 GGTCAACAACAGCTCTTCCAGATGTCGAATACGGCGGATGATGCCGCACTTGCACA 1073
:
: Oy 234 PAspThrSerAspAspAspPhe----- 241
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 1074 GGTCCGCTGAGCAGCAGCAGGATTCGTGAAAGCCGCTTGATGAGCTGCTCATTTGC 1133
:
: Oy 242 -----MetGluGlyGlyGlyGluLysPglGlySe 252
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 1134 GAACGTGCGCTGTGTCAGATCTTTCGCTGATGAGAGCGCGT---GCCGCGCGCGGG 1190
:
: Oy 252 RAspGlyMetGlyLysPglYser----- 260
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 1191 CGGTGAGCTAGGACCATGATGCTCTTTAGACGCGATCACCGGCGCTTACGCCCGCA 1250
:
: Oy 261 -GluPheLeuGlnArgAspPheSerGluTrpGlyuArgTrpHisThrGlnSerLeu-- 279
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 1251 GCTGTCTCTCAACAGAGCTTCTG-----TACCATCGCAACAGTGGCAT 1295
:
: Oy 279 ----- 279
:
: Db 1296 CAAGCCGCTGTCTCACCATCTTCGACCGCATATTGTGAGAGCTGATTAAGATGAA 1355
:
: Oy 280 -----GlnAspMetSerLysGlnGluLeu----- 287
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 1356 GCGCTGAATCTCAGCCAGCGAGCTGTCTGCTTGAAGCCATCATCTACTGTCAACCC 1415
:
: Oy 288 -----IleuGlyGlyTrpLeuGluGlyLysCysLeuSerArgMetGlu-- 302
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 1416 GGACATACGGCGGCAAGACCGCGCGAGATGAGATGCGCGCAAGAGTGTACGC 1475
:
: Oy 303 -----AspGlnAsnAsnAlaGluArgLeuGlnSerLysArgLeuGlyGlyAspAspAl 320
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 1476 TTGCTTGACAGCAGCAGCCGCGCTGGAACATCCG-----GCCGACGATGG 1520
:
: Oy 320 aArgValArgGluLeuGlnGluLeuAspArg-LeuArgAlaGluAsnLeuGlnLeu 340
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 1521 ACGCTTTGCGCACTGCTGCTGCTGCTGCGCGCTTTTCCATTCATGACCTGAAGTCCC 1580
:
: Oy 340 eUrThrGlnAsnGluLeuHisArgGlnGlnAlaArgAlaProLeuSerLys 356
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 1581 AGGATCACCTGTCTCTTCCTTCGCAATTACACAGCAGCGCGCTGAGAGAG 1630
:
: RESULT 9
: US-08-486-403-1
:   Sequence 1, Application US/08486403
:   Patent No. 6281330
:   GENERAL INFORMATION:
:     APPLICANT: EVANS, RONALD M.
:     APPLICANT: MCKEOWN, MICHAEL B.
:     APPLICANT: ORO, ANTHONY E.
:     APPLICANT: SEGRAVES, WILLIAM A.
:     APPLICANT: YAO, TSO-PANG
:     TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE
:     TITLE OF INVENTION: STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE
:     NUMBER OF SEQUENCES: 29
:     CORRESPONDENCE ADDRESS:
:       ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
:       STREET: 444 South Flower Street, Suite 2000
:       CITY: Los Angeles
:       STATE: California
:       COUNTRY: United States
:       ZIP: 90071
:     COMPUTER READABLE FORM:
:       MEDIUM TYPE: Floppy disk
:       COMPUTER: IBM PC compatible
:       OPERATING SYSTEM: PC-DOS/MS-DOS
:       SOFTWARE: PatentIn Release #1.0, Version #1.25
:     CURRENT APPLICATION DATA:
:       APPLICATION NUMBER: US/08/486,403
```

FLING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/907,908
FLING DATE: 02-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P41 9321
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 546-4737
TELEFAX: (619) 546-9392
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2304 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..1701
US-08-486-403-1
Alignment Scores:
Pred. No.: 0.00669 Length: 2304
Score: 141.00 Matches: 106
Percent Similarity: 32.43% Conservative: 49
Best Local Similarity: 22.18% Mismatches: 127
Query Match: 7.38% Indels: 197
Gaps: 22
US-09-972-758a-2 (1-359) x US-08-486-403-1 (1-2304)
QY 13 ProGlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGlnGluLeuAsnProGluArg 32
DB 388 CCCAATTGGCTGGAGGCTCTGCCGCCCTGCAGTCCAGACGACTATCCGCTTAACCAT 447
QY 33 ProProGlyAlaGlnGluArg-----ValProGlnGluAspSerArgTrpGlnSer 49
DB 448 CCGGTGAGCGGACAGACACCTCTGCTCTATTGG-CGGGATGGGGCGAGTGGCAACCA 506
QY 50 ArgAlaPheProGlnLeuGlyGlyArgProGly-----ProGlnGlyGlnGly 65
DB 507 CTACGGCGCTGACAGCTGTGAGGCTGCACAGGCTCTTTAAACGCCACAGTCCCAACGA 566
QY 66 Ser-----LeuGlnSerGlnProProProLeuGlnThrGlnAlaCysProGlnSer 82
DB 567 TCTCACAATACGCTTGCAGAGGAGAACCGCACTGCATCTATACAAAGCGGAGAGGAA--- 623
QY 83 SerCysLeuArgGlnGlnGlyGlnGlnAsnGlyAspSerSerAlaGlyGlyAsp 102
DB 623 ----- 623
QY 103 PheProProAlaGlnValGlnProThrProGlnAlaGlnLeuLeuAlaGlnProCys 122
DB 624 -----CCGCTCCAGTACTGCCGCTACCAAGAGTCCCTAACCTGCGG----- 665
QY 123 HisAspSerGlnAlaSerLysLeuGlyAla-----ProAlaAlaGlnGlyGln 138
DB 666 CATACAGCGGACGCGTCCAGAGGAGAGCGTCAACGCGGCGCCCGCAATGCGGCGG--- 722
QY 139 GlnGlnTrpGlnGlnGlnArg----- 146
DB 723 -----TAGGCTCAGCGCAGCGAGGCGGCGTACGCGTTCGTAAGCGGATC 776
QY 147 GlnLeuGlnLysLysLysLysArgArgProSerLysLysLysArgHisTrpLysPro 166
DB 777 CAGCTTCACAGCGGAGAGGAGGAGCGG-----CGTTCTGCGCGGAT 821
QY 167 TyrTrpLysLeuThrTrpGlnGlnLysLysLysPhe-AspGlnLysGlnSerLeuArgAl 186
DB 822 -----GGGACAGGCGACAGGTTCTGATGACTTTCATGACCAATACGCT 863

QY 186 aserArg-----IleArgAlaGlnMetPheAlaLysGlyGln-- 198
DB 864 GTCCAGGATTTTCTGATCGAGCGCATATAGAGCCAGCAGGAGCGGAGACCAATG 923
QY 199 -----ProValAlaProTyrAsnThrGlnPheLeu 210
DB 924 CGGCGATGCTGACTGACGCTCTGCGCGTGTGGTTCCTATTCCACAGTCCAG----- 975
QY 210 LaspAspHisAspGlnGlnGlnProAspLeuLysThrClyLeu----- 224
DB 976 -----CCGACTACAAAGGTCCTGTGCGGCTGTGCCAAGT 1013
QY 225 -----TyrSerLysArgAlaAlaAlaLysSerAs 234
DB 1014 GGTCAACAACAGCTCTTCCAGATGTCGTGATACCGCGCATGATGCCGCACTTGGCCA 1073
QY 234 PAspThrSerAspAspPhe----- 241
DB 1074 GGTGCGGCTGGACACACAGTATTCTGTGAAGCCGCTTGATCGAGCTGCTCATTTGC 1133
QY 242 -----MetGlnGlnGlyGlnGlnGlnAspGlyGlySe 252
DB 1134 GAACGTGGCTGTGTGCACCATCGTTTGTGATGACCGCGT---GCCGCGCGGCGG 1190
QY 252 rAspGlyMetGlyGlyAspGlySer----- 260
DB 1191 CGGTGACATACGCGCAGCATGTGCTCTTGACGACGATCACCAGGCGCTTCAGCCCA 1250
QY 261 -GlnPheLeuGlnArgAspPheSerGlnTrpGlyArgTyrHisThrGlnSerLeu-- 279
DB 1251 GCTGTCTCTCAACAGAGCTTCTCG-----TACCATGCGAACACTGCGAT 1295
QY 279 ----- 279
DB 1296 CAAGCCGCTGTGCACCATCTTGCACCGCATTTGTGAGCTGACTGATTAAGATGAA 1335
QY 280 -----GlnAsnMetSerLysGlnGlnLeu----- 287
DB 1356 GCGGCTGATCTCGACCGACGAGAGCTGTCTGTGAAGCCATCATCTGTACAAACC 1415
QY 288 -----IleLysGlnTrpLeuGlnGlnGlnLysCysLeuSerArgMetGln-- 302
DB 1416 GGACATACCGCGGATCAAGAGCCGCGGCGGATGAGATGCGCGCAGAAAGGTGTACGC 1475
QY 303 -----AspGlnAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlyAspAspAl 320
DB 1476 TTGCTGTGACGACGACGCTGCGCTGGAACATCG-----GGCGACGATGG 1520
QY 320 aArgValArgGlnLeuGlnGlnLeuAspArg-LeuArgAlaGlnAsnLeuGlnLeu 340
DB 1521 ACCTTTGCCCACTGCTGCTGCTGCGCGCTTGGCATGATCAGCTGAAGTGGC 1580
QY 340 eutThrGlnAsnGlnLeuHisArgGlnGlnGlnAlaProLeuSerLys 356
DB 1581 AGGATCACTGCTTCTCTTCCGATTTACACGACGAGCGGCGCTGAGGAG 1630
RESULT 10
US-09-265-013-2
Sequence 2, Application US/09265013
Patent No. 6451304
GENERAL INFORMATION:
APPLICANT: FRIEDMAN, Theodore
APPLICANT: MIYANOVARA, Atsushi
TITLE OF INVENTION: METHOD FOR RETROVIRUS VECTOR PRODUCTION BY SEPARATED
TITLE OF INVENTION: GAG AND POL EXPRESSION
FILE REFERENCE: 041673/2010
CURRENT APPLICATION NUMBER: US/09/265,013
NUMBER OF SEQ ID NOS: 1999-03-09
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1617

```

: TYPE: DNA
: ORGANISM: Moloney murine leukemia virus (MOLV)
: US-09-265-013-2

Alignment Scores:
Pred. No.: 0.00899 Length: 1617
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88% Mismatches: 152
Query Match: 7.17% Indels: 147
DB: 4 Gaps: 17

US-09-972-758a-2 (1-359) x US-09-265-013-2 (1-1617)

QY 18 CysThrGly---AlaIaIaIaValGInGluLLeuAsnProGluArpProGluAl 36
   ||||| |||:|||||
Db 300 TGACACCTTAAGCTTCGCCCTCTCTCTCCATCGCGCCGCTCTCTCCCTTGAAAC 359

QY 36 aGluGluArGValProGluGluAspSerArGTrpGInSerArGAlaPheProGluLeuG 56
   ||| |||: |||
Db 360 TCCTCGTTCGACCCCGCTCGATCTCTCCCTTATCCAGCCCTCAGT---CCTTCTCAGG 416

QY 56 yGluArGProGluProGlu-----GlyGluGlySer-----LeuG 68
   ||| |||: |||
Db 417 CGCCAAACCTAAACCTCAAGTCTTCTCTGACAGTGGGGGCCCTCATCGACCTACTTAC 476

QY 68 userGInProProProGluGluThrGInaIaCysProGluSerSerCysLeuArGLeuG 88
   ||||| |||: |||
Db 477 AGAAGACCCCGCTTATAGGAGACCCAGACCCCTTCGAC-----AGGAGCGG 530

QY 88 yGluLyGlyGInAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaG 108
   ||||| |||: |||
Db 531 A-----AATGCTGAGAGAGAGACCCCTGGGGGAGAGACCG----- 567

QY 108 uValGluProThProGluAlaGluLeu-----AlaGInProCysHISas 124
   ||||| |||: |||
Db 568 ----GACCCCTCCCATGACATCTCGCTACCTGGGAGAGGAGACCCCTGTGGCGGA 623

QY 124 pSerGluAlaSerLysLeuGlyAlaProAlaGInaGlyGluGluGluTrpGInG 144
   ||||| |||: |||
Db 624 CTCACCTACTCCGACAGCATTCCTCCCGCAGAGAGGAAAC----- 666

QY 144 nGluArGInLeuGlyLysLysLysHisArGArGArProSerLysLysLysArGHisTr 164
   ||||| |||: |||
Db 667 -----GGACAGCTTCAATACATGCGCGCTTCTCTCTGACACTTACAACTG 713

QY 164 pLysProTyTrpLysLeuThrTrpGluGluLysLysLysPheAspGluLysGInSerLe 184
   ||||| |||: |||
Db 714 GAAA-----AATAATAACCTCTTCTTTCTGAGATCCAGGTAA 752

QY 184 uArGAlaSerArGIlleArGAlaGluMetPheAlaLysGlyGInProValAlaProTyris 204
   ||| |||: |||
Db 753 ACTGACAGCTCTGATGACGTCTTCTCTCATC----- 783

QY 204 nThrThrGInPheLeuMetAspAspHisAspGInGluProAspLeuLysThrGly-- 223
   ||| |||: |||
Db 784 -ACCATCAAGCCACTGGGAGACATGTGACGACGTGGGAGCTCTGTCGACCGAGA 842

QY 224 -----LeuTySerLysArGAlaIaIaLysSerAspAsp----- 235
   ||| |||: |||
Db 843 AGAAAAACAACGGCTCTCTTAGAGGCTAGAAAGGCGTCCGGGGCATGATGAGCCGCC 902

QY 236 -----ThSerAspAspAsp 241
   ||| |||: |||
Db 903 CACTCAACTGCCAATGAAGTCGATCCGCTTTTCCCTTCGAGCCGCCAGCTGGAGATT 962

QY 241 eMetGluGluGlyGlyGluGlu----- 248
   ||| |||: |||
Db 963 CACCAACCCAGGAGGTAGGAACACACTAGTCACATAGCGCAATTGCTCTACGGGTCT 1022

QY 249 -----AspGlyGlySerAspGlyme 255
   ||| |||: |||
Db 1023 CCAAAAGCGCGGAGAGGCCCAACAATTGGCCAAAGGTAAAGCAATTAACACAGCGCC 1082

```

```

QY 255 tGlyGlyAspGlySerGluPheLeuGInaRgasPheSerGluThr-TyrGluArGTrp-- 274
   ||| |||||:|||||
Db 1083 CAATGACTCTCCCTCGGCCCTTCTAGAGAGA---CTTAAGAGAGCTATCGACGTACAC 1139

QY 274 ----- 274
Db 1140 TCCTTATGACCCCTGAGGACCCAGGAGAGAAACTAATGTGTATGCTTCTTATTTGGCA 1199

QY 275 -----HisThrGluSerLeuGInaSerLys 284
   ||| |||||: |||
Db 1200 GTCTGCCCCAGACATTGGAGAAAGTTAGAGACGATTAGAAAGATTAAAAACAGACCT 1259

QY 284 sGInGluLeuIleLysGluTyTrpLeuGInuLysCysLeuSerArGMetGInAspG 304
   ||||| |||: |||
Db 1260 TGGAGATTGTGGTTAGAGAGCGAGAAAGATCTTTAATTAACGAGAAACCCCGAGAGAAAG 1319

QY 304 uAsnAsnArGLeuArGLeuGInuSerLysArGLeuGlyGlyAspAspAlaArGValArG 324
   ||||| |||: |||
Db 1320 ACAGGAACGTATCAGC-----AGAGA 1340

QY 324 uLeuGluLeuGluLeuAspArGLeuArGAlaGInaSnLeuGInLeuThrGInaG 344
   ||| |||: |||
Db 1341 AACGAGAGAAAAAGAGAAACGCCGTAGACAGAGATGAGACAGAAAGAGAAAGAGAGA 1400

QY 344 uLeuHisArGInGInGluGluArGAlaProLeu 354
   ||| |||: |||
Db 1401 TCGTAGGAGCATATGAGAGATGAGCAAGCTA 1431

RESULT 11
: Sequence 3, Application US/09011745
: Patent No. 6165715
: GENERAL INFORMATION:
: APPLICANT: Collins, Mary KT
: APPLICANT: Weis, Robin A
: APPLICANT: Takeuchi, Yasuhiro
: APPLICANT: Cosset, Francois-Loic
: TITLE OF INVENTION: Expression systems
: FILE REFERENCE: 09/011,745
: CURRENT APPLICATION NUMBER: US/09/011,745
: EARLIER FILING DATE: 1998-06-22
: EARLIER APPLICATION NUMBER: PCT/GB96/02061
: EARLIER FILING DATE: 1996-08-23
: EARLIER APPLICATION NUMBER: GB9517263.1
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 3
: LENGTH: 7308
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Portion of
: US-09-011-745-3

Alignment Scores:
Pred. No.: 0.0723 Length: 7308
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88% Mismatches: 152
Query Match: 7.17% Indels: 147
DB: 4 Gaps: 17

US-09-972-758a-2 (1-359) x US-09-011-745-3 (1-7308)

QY 18 CysThrGly---AlaIaIaIaValGInGluLLeuAsnProGluArpProGluAl 36
   ||||| |||:|||||
Db 1866 TGACACCTTAAGCTTCGCCCTCTCTCTCCATCGCGCCGCTCTCTCCCTTGAAAC 1925

QY 36 aGluGluArGValProGluGluAspSerArGTrpGInSerArGAlaPheProGluLeuG 56
   ||| |||: |||

```

```

Db 1926 TCCTGCTGACCCCGCTGATCTCCCTTATCCAGCCCTCACT---CCTTCTAGG 1982
QY 56 YGLATGProGlyProGlu-----GlyGluGlySer-----LeuG1 68
| : : : : : | : : : : |
Db 1983 CGCCAAACCTAAACCTCACTTCTTCTGACAGTGGGGGCGCTCATGCACTACTTAC 2042
QY 68 uSerGlnProProLeuGlnInThrGlnAlaCysProGluSerSerCysLeuArgGluG1 88
| : : : : : | : : : : |
Db 2043 AGAAGACCCCGCTTATAGGACCAAGACCCCGCTTCCGAC-----AGGACGG 2096
QY 88 YGluLYsGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaG1 108
| : : : : : | : : : : |
Db 2097 A-----AATGTGGAAGAGCGACCTCGGGAGGACCG----- 2133
QY 108 uValGluProThrProGluAlaGluLeu-----AlaGlnProCysHis 124
| : : : : : | : : : : |
Db 2134 ----GACCCCTCCCAATGCACTCTCGCTACGCTGGAGACGGAGCCCGCTGTGGCGCA 2189
QY 124 pSerGluAlaSerLysLysLysGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnG1 144
| : : : : : | : : : : |
Db 2190 CTCACACTCTCCGACGATTCCTCCCTCCGCGAGAGGAAC----- 2232
QY 144 ngInArGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysArgHisTr 164
| : : : : : | : : : : |
Db 2233 -----GGACAGCTTCAATACGCGCTTCTCTCTGACCTTACACTG 2279
QY 164 pLysProTyrTyrLysLeuThrTrpGluGluLysLysPheAspGluLysGlnSerLe 184
| : : : : : | : : : : |
Db 2280 GAAA-----AATATACCTCTCTTTTCTGACATCCAGGTAA 2318
QY 184 uAlaGlnAlaSerAlaIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAs 204
| : : : : : | : : : : |
Db 2319 ACTGACAGCTCTGATGAGTCTGCTCTCATC----- 2349
QY 204 nThrTrGlnPheLeuMetLaspAspHisAspGlnGluGluProAspLeuLysThrGly-- 223
| : : : : : | : : : : |
Db 2350 -ACCATCAAGCCCACTGGGACACCTGTCAAGAGCTTGGGAGCTCTGCTGACCGAGA 2408
QY 224 -----LeuTyrSerLysArgAlaAlaAlaLysSerAspS-- 235
| : : : : : | : : : : |
Db 2409 AGAAACCAACAGGGGTGCTTACAGGCTAGAAAGCGGTCGCGGGGATGATGGGCGCC 2468
QY 236 -----ThSerAspAspSph 241
| : : : : : | : : : : |
Db 2469 CACTCACTGCCCAATGAATGATGATGCGCTTCTCCCTGACGCGCCACACTGGGATTA 2528
QY 241 eMetGluGluGlyGluGlu----- 248
| : : : : : | : : : : |
Db 2529 CACCACCCAGGAGGTAGGACCACTAGTCACATCGCACTTGTCTTACGGGTCT 2588
QY 249 -----AspGlyGlySerAspGlyMe 255
| : : : : : | : : : : |
Db 2589 CCAAACGGGGGAGAGACCCCAATTTGGCCAAAGGTAAAGGATATACACAGAGGCC 2648
QY 255 tGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTr-- 274
| : : : : : | : : : : |
Db 2649 CAATGAGTCTCCCTCGGCTTCTTCTAGAGAGA---CTTAAGAGAGCTATCGCAGTACAC 2705
QY 274 ----- 274
| : : : : : | : : : : |
Db 2706 TCCTTATGACCTTGAGGACCCAGGCAAGAACTAATGTCTATGTCTTTCATTGGCA 2765
QY 275 -----HisThrGluSerLeuGlnAsnMetSerL 284
| : : : : : | : : : : |
Db 2766 GTCTGCCCCAGACATTGGGAGAAATTAGAGAGTTAGAAAGATTAAACAAAGAGCT 2825
QY 284 sGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetClnuSpG1 304
| : : : : : | : : : : |
Db 2826 TGGAGATTGGTTAGAGAGCAGAAAGATCTTATTAACGAGAAACCCCGGAGAAAG 2885
QY 304 uAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAlaArgAlaArgG1 324
| : : : : : | : : : : |
Db 2886 ACAGGAACGTATCAG-----AGAGA 2906

```

```

QY 324 uLeuGluLeuGluLeuAspArgLeuArgAlaGlnAsnLeuGlnLeuThrGluAsnG1 344
| : : : : : | : : : : |
Db 2907 AACAGACGAAAGAAAGAGAGAGCGCTTACAGAGAGATGACAGAAAGAGAGAGAG 2966
QY 344 uLeuHisArgGlnGlnGluArgAlaProLeu 354
| : : : : : | : : : : |
Db 2967 TCCTAGACAGCATATGACAGATGACAGAAAGCTA 2997

RESULT 12
US-09-011-745-4
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Lolc
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; US-09-011-745-4

Alignment Scores:
Pred. No.: 0.0723 Length: 7308
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88% Mismatches: 152
Query Match: 7.17% Indels: 147
Gaps: 17

US-09-972-758a-2 (1-359) x US-09-011-745-4 (1-7308)
QY 18 CysThrGly---AlaAlaAlaValGlnGluGluLeuAsnProGluArgProPro-GlyAl 36
| : : : : : | : : : : |
Db 1866 TGTACACCTTAAGCCCTCCGCTCTCTCTCCATCCGCGCTCTCCCTTGAACC 1925
QY 36 aGluGluArgValProGluGluLysSerArgTrpGlnSerArgAlaPheProGlnLeuG1 56
| : : : : : | : : : : |
Db 1926 TCCTGCTGACCCCGCTCGATCTCTCTTATCCAGCCCTCACT---CCTTCTTAA 1982
QY 56 YGLATGProGlyProGlu-----GlyGluGlySer-----LeuG1 68
| : : : : : | : : : : |
Db 1983 CGCCAAACCTAAACCTCAAGTCTTCTGACAGTGGGGGCGCTCATGCACTACTTAC 2042
QY 68 uSerGlnProProProLeuGlnInThrGlnAlaCysProGluSerSerCysLeuArgGluG1 88
| : : : : : | : : : : |
Db 2043 AGAAGACCCCGCTTATAGGACCAAGACCCCGCTTCCGAC-----AGGACGG 2096
QY 88 YGluLYsGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaG1 108
| : : : : : | : : : : |
Db 2097 A-----AATGTGGAAGAGCGACCTCGGGAGGACCG----- 2133
QY 108 uValGluProThrProGluAlaGluLeu-----AlaGlnProCysHis 124
| : : : : : | : : : : |
Db 2134 ----GACCCCTCCCAATGCACTCTCGCTACGCTGGAGACGGAGACCCCGCTGTGGCGCA 2189
QY 124 pSerGluAlaSerLysLysLysGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnG1 144
| : : : : : | : : : : |

```

```
Db 2190 CTCACACTCTCGGAGGACATTCCTCCCGGAGGAGAAAC----- 2232
Qy 144 nglarglnleucllyslslyshisargargproserlyslslysrghisr 164
Db 2233 -----GGACAGCTTCAATACTAGCCGCTTCCTCTTGTGACTTTACACTG 2279
Qy 164 plysprotyrtyrlyslsleuthrtpglnlulslslyshsphaaspglulysglnserle 184
Db 2280 GAAA-----AATAATACCCCTTCTTCTGTAAGATCCAGCTAA 2318
Qy 184 uargalaserarglleargalaglmelephalaalysgllylnprovalalaprotyras 204
Db 2319 ACTGACACTCTGTGATCGAGTCTGTTTCATC----- 2349
Qy 204 nThrThrglnpheleumetaspaasphiaspqlnglnluproaspleuylsrhgly-- 223
Db 2350 -ACCATTCAGCCCACTGGAGCGACTGTACAGACTGTGGGAGACTGTGCTGACCGGAGA 2408
Qy 224 -----LeuTyrserysargalalaalalysserasasp----- 235
Db 2409 AGAAACACACGGGCTCTCTTAGAGCGTAGAAGCGGTGCGGGCGCATGTATGGGCCCCC 2468
Qy 236 -----Thrserysaspasp 241
Db 2469 CACTCACTGCCAATGAATGATGCCGCTTTTCCCTCGAGCGCCGACACTGGAGATTA 2528
Qy 241 emetglnlulgllygllyglu----- 248
Db 2529 CACCACCCAGGAGAGAGCCCAACCACTAGTCCACTATGCCAGTTGCTCTTACGGGGTCT 2588
Qy 249 -----Aspqllylseryaspqlyme 255
Db 2589 CCAAAACGCGGCGGAGAACCCCAATTTGGCCAGTAAGTAAGAAATACACAAGGCC 2648
Qy 255 tglglyaspqlyseryglupheleuglnargaspsererlyuhtrtyrglnargtyr-- 274
Db 2649 CAATGAGTCTCCCTGGCTTCCTAGAGAGA--CTTAAGGAGCCATTCACAGGTACAC 2705
Qy 274 ----- 274
Db 2706 TCCATTATGACCTTAGAGGACCCAGGCAAGAACTAATGTCTATGTCTTTCATTGGCA 2765
Qy 275 -----Histhrclyserleuglnasmetserly 284
Db 2766 GTCTGCCCCAGACATTGGAGAAAGTTAGACAGCTTAGACATTAAACACAGACCT 2825
Qy 284 sglnglnleucllyslslyshisargargproserlyslslysrghisr 304
Db 2826 TGGAGATTGGTTAGAGAGGAGCAAAAGATCTTTAATAACGAGAAACCCCGGAGAAAG 2885
Qy 304 uasnAsnArgleuArgleuglnserlysrghisrlygllyaspaspalaargvalarg 324
Db 2886 AGAGGAAACGTATCAG-----ACGGA 2906
Qy 324 uleucllyslslyshisargargalaglulslsleuglnleucllysrghisr 344
Db 2907 AACAGAGGAAACAGACCCCTAGGACAGAGATGACAGAAAGAGAAAGAGAA 2966
Qy 344 uLeuHlsArgGlnGlnArgAlaProLeu 354
Db 2967 TCCTAGGAGACATAGAGATGACCAAGCTA 2997
```

```
RESULT 13
US-09-011-745-2
: Sequence 2, Application US/09011745
: Patent No. 6165715
: GENERAL INFORMATION:
: APPLICANT: Collins, Mary KL
: APPLICANT: Weiss, Robin A
: APPLICANT: Takeuchi, Yasuhiko
: APPLICANT: Cosset, Francois-Lolc
: TITLE OF INVENTION: Expression systems
: FILE REFERENCE: 09/011,745

: CURRENT APPLICATION NUMBER: US/09/011,745
: CURRENT FILING DATE: 1998-06-22
: EARLIER APPLICATION NUMBER: PCT/GB96/02061
: EARLIER FILING DATE: 1996-08-23
: EARLIER APPLICATION NUMBER: GB9517263.1
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 7616
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-2

Alignment Scores:
Pred. No.: 0.0766 Length: 7616
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88% Mismatches: 152
Query Match: 7.17% Indels: 147
DB: Gaps: 17

US-09-972-758a-2 (1-359) x US-09-011-745-2 (1-7616)

Qy 18 CysThrGly---AlaAlaAlaValGlnGlnLeuAsnProGlnuArgProPro-GlyAl 36
Db 1027 TGTACACCTTAAGCTCTCCCTCTCTTCCTCCATCCGCGCTCTCTCTTGAACC 1086
Qy 36 aglnluargvalprogluinspserfargtrpqlnserargalapherproglu 56
Db 1087 TCCTCGTTGACCCCGCTGCATCTCTCTTATTCACCTCACT---CCTTCTTAGG 1143
Qy 56 yglYargProGlyProGlu-----GlyGlnGlyser-----LeuG 68
Db 1144 CGCCAAACCTTAACCTCAAGTTCTTTCAGACAGTGGGGGCGCTCATCGACTACTTAC 1203
Qy 68 uSerGlnProProProleuGlnThrGlnAlaCysProGlnuSerSerCysLeuArgGln 88
Db 1204 AGAAGACCCCGCTTATAGGAGCCAGACCCCTTCCGAC-----AGGACGG 1257
Qy 88 yglulysgllyGlnsnGlyaspaspserSerAlaGlyGlyaspheProProProAla 108
Db 1258 A-----AATGTTGAGAGAGGACCCCTGCGAGAGCACC----- 1294
Qy 108 uValGlnProThrProGlnuAlaGlnLeu-----AlaGlnProCysHis 124
Db 1295 -----GACCCCTCCCAATGCGATCTCGCTAGGAGAGGAGGAGCCCTGTGGCCGA 1350
Qy 124 pSerGlnAlaSerlyslsleuglnalaproalalagllygllygluGlnluprpglly 144
Db 1351 CTCACACTCTCGGAGGACATTCCTCCCGGAGGAGAAAC----- 1393
Qy 144 nglarglnleucllyslslyshisargargproserlyslslysrghisr 164
Db 1394 -----GGACAGCTTCAATACTAGCCGCTTCCTCTTGTGACTTTACACTG 1440
Qy 164 plysprotyrtyrlyslsleuthrtpglnlulslslyshsphaaspglulysglnserle 184
Db 1441 GAAA-----AATAATACCCCTTCTTCTGTAAGATCCAGCTAA 1475
Qy 184 uargalaserarglleargalaglmelephalaalysgllylnprovalalaprotyras 204
Db 1480 ACTGACACTCTGTGATCGAGTCTGTTTCATC----- 1510
Qy 204 nThrThrglnpheleumetaspaasphiaspqlnglnluproaspleuylsrhgly-- 223
Db 1511 -ACCATTCAGCCCACTGGAGCGACTGTACAGACTGTGGGAGACTGTGCTGACCGGAGA 1569
Qy 224 -----LeuTyrserysargalalaalalysserasasp----- 235
```


QY 236 -----ThrsrAspaspaph 241
Db 6244 CACTCAACTGCCCATGAATGCATGCCGCTTTCCCTCGAGGCCAGACTGGGATTA 6185
QY 241 eMeGluGluGlyGlyGluGlu----- 248
Db 6184 CACCACCCAGCAGGTAGGAACACACTAGTCCACTATCGCAGTTGCTCTAGCGGCTCT 6125
QY 249 -----AspGlySerAspGlyme 255
Db 6124 CCAAAACGGCGGAGAACCCCAACCAATTGGCCAGGTAAGAGAAATACACAAAGGCC 6065
QY 255 fGlyGlyAspGlySerGluPheGluArgAspPheSerGluThrGlyArgGly-- 274
Db 6064 CAATGAGTCTCCCTCGGCTCTCTAGAGAGA---CTTAAGAACCCCTATCGCAGCTTAC 6008
QY 274 ----- 274
Db 6007 TCCTTATGACCCTGAGGACCCAGGCAAAACTAATGTCTATGCTTTTCATTGGCA 5948
QY 275 -----HisthGluSerLeuGlnAsnMetSerLy 284
Db 5947 GTCTGCCCCACACATTGGCAGAAAGTTAGAGAGCTTGAACATTTAAAGAAAGACGCT 5888
QY 284 sGluGluLeuLLeuGlyLeuGluLeuGluLysCysLeuSerArgMetGluAspG1 304
Db 5887 TGCAGATTGTGTTAGAGAGCGCAAAAGATCTTAATAACGAGAAACCCCGGAGAAAG 5828
QY 304 uAsnAsnArgLeuArgLeuGluSerLyArgLeuGlyGlyAspAspAlaArgValArgG1 324
Db 5827 AGAGGAACGTATACAG-----AGAGA 5807
QY 324 uLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuThrGluAsnG1 344
Db 5806 AACGAGAGAAAAGAAAGAACCCCTAGCAGAGAGATGACGAGAAGAAAGAAAGAGA 5747
QY 344 uLeuHisArgGlnGlnGluArgAlaProLeu 354
Db 5746 TCGTAGGACATAGAGATGACGAGCTA 5716

RESULT 15
US-08-850-961-1
Sequence 1, Application US/08850961
Patent No. 6013517
GENERAL INFORMATION:
APPLICANT: Respass, James G.
APPLICANT: De Polo, Nicholas J.
APPLICANT: Chada, Sunil
APPLICANT: Sauter, Sybille
APPLICANT: Bodner, Mordechai
APPLICANT: Driver, David A.
TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation, Intellectual Property - R440
STREET: P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/850,961
FILING DATE: 05-MAY-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kruse, No. 6013517man J.
REGISTRATION NUMBER: 35,235
REFERENCE/DOCKET NUMBER: 930049, 424C4 / 1147,005

TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3520
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8332 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-850-961-1
Alignment Scores:
Pred. No.: 0.0867 Length: 8332
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Mismatch: 42
Best Local Similarity: 20.88% Indels: 147
Query Match: 7.17% Gaps: 17
US-09-972-758A-2 (1-359) x US-08-850-961-1 (1-8332)
QY 18 CysThrGly---AlaAlaAlaValGlnGluLeuAsnProGluArgProPro-GlyAl 36
Db 920 TGTACACCTTAAGCCTCCGCTCTCTCTCCATCCGCCCGCTCTCCCTGGAAC 979
QY 36 aGluGluArgValProGluLysPheArgGlyProGlnSerArgAlaPheProGlnLeuG1 56
Db 980 TCTCTGTCGACCCCGCTCGATCTCTCTTTATCCAGCCCTACT---CCTTCTCTAGG 1036
QY 56 yGlyArgProGlyProGlu-----GlyGluGlySer-----LeuG1 68
Db 1037 CGCCAAACCTAAACCTCAAGTCTTCTGACAGTGGGGCGCGCTCATGACCTACTTAC 1095
QY 68 uSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluG1 88
Db 1097 AGAAGACCCCGCGCTTATAGGAGCCCAAGACCAACCCCTTCGAC-----AGGACGG 1150
QY 88 yGluLysGlyGlnAsnGlnAspAspSerSerAlaGlyGlyAspPheProProAlaG1 108
Db 1151 A-----AATGTGAGAGAGCGACCCCTCGGAGAGACCG----- 1187
QY 108 uValGluProThrProGluAlaGluLeuLeu-----AlaGlnProCysHisAs 124
Db 1188 ---GACCCCTCCCAATGCGATCTCGCTACGTGGAGACGGAGCGCCCTGTGGCGCA 1243
QY 124 pSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluThrProGlyGlnG1 144
Db 1244 CTCACACTCCGCGAGCATTCCTCCCGCGAGGAGAAAC----- 1286
QY 144 nGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisStr 164
Db 1287 -----GGACAGCTTCATATGCTGCGGCTTCTCTCTTCTGACCTTACACTG 1333
QY 164 pLysProTyTrpLysLeuThrTrpGluGluLysLysPheAspGluLysGlnSerLe 184
Db 1334 GAAA-----AATATTAACCTTTCTTTCTGAGATCCAGTAA 1372
QY 184 uArgAlaSerArgLLeuArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyTras 204
Db 1373 ACTGACAGCTCTGATCGACCTCTTCATC----- 1403
QY 204 nThrThrGlnPheLeuMetAspAspHisAspArgGlnGluProAspLeuLysThrGly-- 223
Db 1404 -ACCATCAGCCAGCTCGGAGCAGCTTCAGCAGCTTTGGGAGCTCTGTCGACCGGAGA 1462
QY 224 -----LeuTySerLysArgAlaAlaLysSerAspAsp----- 235
Db 1463 AGAAAGAACAGGGGTCTTGAAGGCTAGAAAGCGCTCGGCGCATGATGCGCGCC 1522
QY 236 -----ThrsrAspaspaph 241
Db 1523 CACTCAACTGCCCATGAATGCATGCCGCTTTCCCTCGAGGCCAGACTGGGATTA 1582

Db	6	PEGE6SLIESPPeLQZQACPESSCLREGBKQNGDSSAGGDPFPPAPEVEPTPEALLAQ	120
QY	121	PCHDSEASKIGAPAAGCEEMGQOOROLGKKHRRRPSKKRRHKPYIKLTWEKKKFE	180
Db	121	PCHDSEASKIGAPAAGCEEMGQOOROLGKKHRRRPSKKRRHKPYIKLTWEKKKFE	180
QY	181	KOSLRASRIAEKFAFGQVAPYNTTQFLMDHDDEEDLTGILYSKRAAKSDTSDDD	240
Db	181	KOSLRASRIAEKFAFGQVAPYNTTQFLMDHDDEEDLTGILYSKRAAKSDTSDDD	240
QY	241	FMEEGGEEDGSDGMGDSFEFLQDFSETEERYHTESLQNMKSQOELIKELYLEKCLSR	300
Db	241	FMEEGGEEDGSDGMGDSFEFLQDFSETEERYHTESLQNMKSQOELIKELYLEKCLSR	300
QY	301	MEENNRRLRESKRIAGDDARVRELLELDRIARAANIQLLENELHROQEAAPLSKGD	359
Db	301	MEENNRRLRESKRIAGDDARVRELLELDRIARAANIQLLENELHROQEAAPLSKGD	359

RESULT 2			
08R409			
ID	08R409	PRELIMINARY:	PTT: 356 AA.
AC	08R409:		
DT	01-JUN-2002 (TrEMBLrel. 21, Created)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Cardiac lineage protein 1.		
CN	Clp1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eultheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129/SVJ;		
RA	Huang F., Wagner M., Siddiqui M.;		
RT	"Structure, expression, and functional characterization of the mouse		
RT	Clp-1 gene."		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
SR	EMBL: AY090614; AM09026.1; --		
SO	SEQUENCE 356 AA; 40243 MW; 242DE7EE66BA293B CRC64;		

Query Match	83.1%	Score 1587.5	DB 11	Length 356
Best Local Similarity	85.8%	Pred. No. 1.7e-105		
Matches 308, Conservative	11	Mismatches 37	Indels 3	Gaps 2

QY	I	MAEPFLSEYHOQOTSCTGTAAGAAYOEBLNERPPGAEEKVPEDSDSWOSRAPFOLCGRPG	60
Db	1	MAEFLLETHOHQOTSNCGTGAAYVHEHSTSERPSPAEEPKDSSWOSRASLOSGRPG	60
QY	61	PEGGSLSIESOPPLLOTACPESSCLTBEGEKGGONGDDSSAGGDPPPAPAEVPTPEALLAO	12
Db	61	QEBBGSLIKHLNLPPLQTNACPGLSLEKEKGQGQGBDLSTGG-ASPSAEEPMSES--LVQ	11
QY	121	PCHDSSEASKLGAPAAAGEEMEQOOQLGKKKKRRRRSKKKRKWKPYRYKLTMEEKKFE	18
Db	118	PGHDSSETTKGEARPAAGSEBPWGQQQLGKKKKRRRRSKKKRKWKYYLYLTMEKKKFE	17
QY	181	KOSTLRASRIIAEMFAKQPAPARYNTTQFLMDDHDQEEDPLKTGLYSKRAAKASDDTSDD	24
Db	178	KOSTLRASRYRAEMFAKGPAPARYNTTQFLMDDHDDEEDPLKTGLYPKRAAKASDDTSDD	23
QY	241	FMEEGGEDDGSGMGGDGSEFIQRDPSEFYERYHTESIQNNSKOPLIKEYILELEKCLR	30
Db	238	FVEBAGEEDDGSGMGGDGSEFIQRDPSEFYERYHAESIQNNMSKOELIKEYILELEKCLR	29
QY	301	MEENNRILRESKRILIGDDARVELTELDELDRIPAEMLQILTLENLHQDERAPLSFGD	359
Db	298	KEDENNRLRESKRILIGGVDAHYVELTELDRLIPAENLQILTLENLHQDERAPLSFGD	356

RESULT 3
Q8VDF7

ID	Q8VD67	PRELIMINARY:	PRT:	246 AA.
AC	Q8VD67			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Similar to HMBA-Inducible.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC022111; AAH2111.1;			
SO	SEQUENCE 246 AA; 28572 MW; 6CF8227F599B77EF CRC64;			
Query Match	62.8%	Score 1199;	DB 11;	Length 246;
Best Local Similarity	93.5%;	Pred. No. 5.4e-78;		
Matches 229;	Conservative 5;	Mismatches 11;	Indels 0;	Gaps 0;

QY	115	AELLAPCHHSESKJGAPACGEEEMGOOLGKKHRRRPSKKKNKPYUKLME	174
	:		
Db	2	SESLVGRGHDSVTKCEAPRACGEEEMGOOLGKKHRRRPSKKKNKPYUKLME	61
QY	175	KKKFKDEKOSLRASRIAREMFAKGOVPARYNTTQFLMDDHDOEERDJKTGLSKRAAKSD	234
	:		
Db	62	KKKFKDEKOSLRASRIAREMFAKGOVPARYNTTQFLMDDHDOEERDJKTGLSKRAAKSD	124
QY	235	DTSDDDFMEBGGDEEDGSDGMDGDSSEFLQRFSETRYHNTSELONNSSKOLKEVLEL	294
	:		
Db	122	DTSDEDFVEAGDEEDGSDGMDGDSSEFLQRFSETRYHNTSELONNSSKOLKEVLEL	184
QY	295	EKCLSRMEDENNRLRLSESKRLGCGDARVARELLELDRLRAENLOLLENELEHROOEAPRL	354
	:		
Db	182	EKCLSRKREDENNRLRLSESKRLGCGDARVARELLELDRLRAENLOLLENELEHROOEAPRL	241
QY	355	SKFGD 359	
	:		
Db	242	SKFGD 246	

RESULT 4	
096MH2	
ID	096MH2
AC	096MH2, PRELIMINARY;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	CDNA FLJ32384 f1s, clone SKM51000104, weakly similar to Homo sapiens
DE	HEXIM1 protein (Similar to putative).
OS	Homo sapiens (Human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
ON	[1]

RC TISSD=SKELTAL MUSCLE:
RA Oshima K., Takahashi-Fuji A., Tanase T., Imose N., Tak
RA Arita M., Mwashino K., Yuuki H., Hata H., Sugiyama T.,
RA Otsuki T., Sato H., Wakamatsu A., Ishi S., Yamamoto J.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamash
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzu
RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases
RN [2]
RP SEQUENCE FROM N.A.
RP TISSUE=EYE:
RC Strausberg R.;
RA Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases
RR EMBL; AK056946; BAB71319.1; -.

Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases
EMBL; AK056946; BAB71319.1; -
DR

DR	EMBL: BC025970.1: -	28.6%	Score 545.5:	DB 4:	Length 286:
SEQUENCE	286 AA: 32418 MW: 56576072096A8A6F CRC64:				
Query Match		43.8%	Pred. No. 2.5e-31:		
Best Local Similarity		36:	Mismatches 93:	Indels 35:	Gaps 6:
Matches 128:	Conservative				
OY	73 PLQIQACFESSCLKEGKQNGDDSS-----SAGCGFPFAVEYEPPEVLELLAOPC--	122			
DB	5 PNQFACNAESPVALEEAFTSCAPSPQTPPERHDSGGSLPTTPRMESHSEDDDLGAVGG	64			
OY	123 --HSEASKLAPAAAGCEEMGQOOROLGKKRKHRRPRPKKRRHMKPYKYLWEEKKRFE	180			
DB	65 LGMMSRSPRRGSPGCSASAE-----AVLARKKHRRRPSKKRRHMRPYLELSMAEQQDE	118			
OY	181 KQSLRASFIRADMEFAKGGQVAPRYNTTQFLMDHDOEEP--DLTKGLYSKRAAASDDTSD	238			
DB	119 ROSORASRVRREMFKGGQVAPRYNTTQFLMDRDPREEPNLDVPHGISHPGSGSES----	173			
OY	239 DDFMEGCEGGGSGMGCGDSEFLQRFPSFETRYEYTHHSLONNKSKOLLEIELEKCL	298			
DB	174 -----EAGDSNGRRRAEGFQKRKPSFETRYEYTHHSLONNKSKOLVADYLELEKRL	224			
OY	299 SRMEDENRRLRESKRRLGGDDAR--VRELELEDLRLAENLQLLTLENELHROO	349			
DB	225 SQAEETFRRLDLOLACTGQSGCROVEELAEVQRLTENORLROENQMWNR	276			
RESULT 5					
09DAC7	PRELIMINARY:	PRT:	280 AA.		
ID	09DAC7:				
AC	09DAC7:				
DT	01-JUN-2001 (TREMBLrel. 17, Created)				
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	4933402L2IRik protein.				
GN	4933402L2IRik.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_Taxid=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Atzawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,				
RA	Saito T., Okazaki Y., Cojibori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Pletschmann W., Gaasterland T., Gissi C., King B., Kochiya H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schriml P.M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hotmann M., Hume D.A., Kaniya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyooka K., Wang C.H., Weitz C., Whitlaker C., Wilming L.,				
RA	Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,				
RA	Yashizaki Y.;				
RT	"Functional annotation of a full-length mouse cDNA collection.";				
RL	Nature 409:665-660(2001).				
DR	EMBL: AK016624; BABS0344.1; -				
DR	MGD: MGI:1918309; 4933402L2IRik.				
SO	SEQUENCE 280 AA: 32388 MW: 68A925FEFA1A63234 CRC64:				

Qy	122	CHSEVSK	-----LGA-----PAAGFEEMG000ROLGKKKHHRRRSKKRH	163
		11	11	11
Db	12	CHDRRSNORMKGLSP	LAMAYVTVTGVEPRAGQAVRW----	RRCMERSTVGAIEAQRH 67
Qy	164	WRKYRVLTMEEKKKPKPEKOSILAS	SLTRAFEMFAGQVAPVNTPTQFLMDDHOEPDLKTG	223
		11	11	11
Db	68	WRPYLELSMAEK00RDEKOSAS	VSAREMPAKGQPLAPVNTTQFLMDDHLEPNI----	124
Qy	224	LYSKRAAAKSDPTSDDDPFMEEGE	-EDGSDQMGDGSGSEFLQRPFSFETERYHYTESLQNM	282
		11	11	11
Db	125	-----DVLHGSHSGSGENMD	ADSDQOGAGHGEFQQRDPFSATERYHTESLQGR	174
Qy	283	SKOELLKEYLELEKCLSRMEDENN	RLRLEKRLGGDAR-----VRELEELDLRAENL	337
		11	11	11
Db	175	SKOELLVDRYLDLERRLSQAE	ETFRRLR-----QLQCCSSRQCGQVEELAAVEERLRTNG	230
Qy	338	QLLTENELHRQ	349	
		11	11	11
Db	231	RLROENEMNRE	242	
RESULT 6				
QyVFN2				
AC	QyVFN2	PRELIMINARY:	PRT:	349 AA.
DT	01-MAY-2000 (TREMBLrel. 13,			Created)
DT	01-MAY-2000 (TREMBLrel. 13,			Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19,			Last annotation update)
DE	CG3508 protein (LD30520P).			
CN	CG3508.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OX	Ephydroidea; Drosophilidae; Drosophila.			
NCBI_TaxID=7227;				
KN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RC	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sulten G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazei R.G., Champe M., Pfeiffer B.D.,			
RA	Man K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,			
RA	Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolintsov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brooksstein P., Brotlier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delber A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Doonan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,			
RA	Glodjer A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibbagam C.,			
RA	Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Krevitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li Q., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy D., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paetle J.M.,			
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Ringden K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,			
RA	Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wassatman D.A., Weinstein G.M., Weissendenbach J.,			
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Abmayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Garlin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacלב J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003705; AAF55085.1; -
DR EMBL: AY051786; AAK93210.1; -
DR FlyBase: FBgn0038251; CG3508.
SQ SEQUENCE 349 AA: 38058 MW: 8597FC7AE08D0747 CRC64;

Query Match 10.9%; Score 209; DB 5; Length 349;
Best Local Similarity 25.9%; Pred. No. 3.1e-07;
Matches 68; Conservative 43; Mismatches 84; Indels 68; Gaps 10;

QY 85 LREGEKQNGD-----DSSAGDFPPPAEVEPTPEAELLAQPCHESEASKLGAAPAGEE 139
DB 1 MAEAVNNESSGQRPDLDSGGGG-----ASGGGVAVGGGS 36
QY 140 EMGQOQROLGKKRRRPSKKRRMKRYKLTWE---EKKRPEKOSLRASRTAEMFA 195
DB 37 GMRKRHRGRKKS-KMOPKTKNHY-PQKLDMSGTAGATLEGNQONSKTKLYRSL- 93
QY 196 KGQPVAPYNTQFLMDHDOEERDLKTYGLSKRAAKSDPTSDDEMEEGEGEDGSDGM 255
DB 94 ---LVYNTNRFLEEMHSE-----LHKDDSDNCF---GSQTE----- 126
QY 256 GGGSEFLQDESETERHTESLQNMKSQELKEYLEKLSRMEDENNRRLRLSKRL 315
DB 127 --DQVFLSKERSDVERARLERLETMSKOELTQECQMLDRLYSKAONISKEP---GAKL 181
QY 316 GGDARVRELELELDRLRAENLQ 338
DB 182 RAQDDKIRQLRSRENQFLRTHLLR 204

RESULT 7
Q8T6B4 PRELIMINARY; PRT; 1142 AA.
AC Q8T6B4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Non-transporter ABC protein AbcF4.
GN ABCF4.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Eukaryota; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4;
RA Anjard C., Loomis W.F.;
RT "Evolution of the ABC transporters of *Dictyostelium*";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF479256; AAL87694.1; -
SQ SEQUENCE 1142 AA: 130224 MW: 33B1815AB09942DC CRC64;

Query Match 7.9%; Score 150; DB 5; Length 1142;
Best Local Similarity 19.8%; Pred. No. 0.02;
Matches 79; Conservative 73; Mismatches 132; Indels 114; Gaps 18;

QY 30 PERPPAEERVPEDSRMOSRAFPQ---LGRPGEGESLESQ-----PPPIQOTA 78
DB 132 PQKGGKQGO--QODSDQDEELIPQPVKGGKAPQKGGKQODSDDEDELIPQPVKGG 189
QY 79 CPSSSCLREGEKQNGDSSAGGDFPPPA-----EVEPTPEAELLAQPC 122

DB 190 KPAPO--KKGGKQOESEDEDEDEVOQPVKGGKNDKKKGVKHVEEHEEHEEIEQPV 247
QY 123 HDEASKLGAAPAG-----EEENGQOQOLGKKHRRRPSKKRRMKRYKLTWE 173
DB 248 --KGGKAPKPKKGGKQOESEDEDDVOQPVKGGKDKKKKHV-----E 294
QY 174 EKKRPEKOSLR-----ASRIAEEMFAKQOPAPYNTQFLMDHDOEERDLKTYGL- 224
DB 295 EEEHEEHEEIEQPVKGGKNDKKKGVKHVE-----EEHEEHEEIEQPVKGG 346
QY 225 ---YSKRAAKSDPTSDDEME-----EGGEEDGSDGMGSDSEFLQDRFSETYER 273
DB 347 SNKKDQKGGKQOQOESEDEDEEIEQPVKGGKDKKKK-----GSKHVEEHEEHEE 400
QY 274 YHTESLQNMKSQELKEYLEKLSRMEDENNRRLRLSKRLGGDDARVRELE- 328
DB 401 EEIE-----QPVKGGKKD-KSLLEDSMSLETSK-KGKKHVEEHEEHEE 450
QY 329 -----LDRRAENQLLTENELHROERAPLSK 356
DB 451 EKKRKSNNKKDKKKKHVE--EEHEEHEEHEEKKPSK 486

RESULT 8
Q9NTH6 PRELIMINARY; PRT; 992 AA.
AC Q9NTH6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Hypothetical 112.6 kDa protein (Fragment).
GN DKEFZP43402413.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Koehler K., Meyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 992 AA: 112628 MW: 039DF5B1E97EF02F CRC64;

Query Match 7.8%; Score 148.5; DB 4; Length 992;
Best Local Similarity 21.1%; Pred. No. 0.022;
Matches 94; Conservative 66; Mismatches 149; Indels 137; Gaps 17;

QY 25 QEELNPERPPGAEEVP-----EEDSRMOSRAF-----POLGPRGEGE----- 64
DB 230 QDELQKQKGLERLSPPLPHERRAOSPPRLATEEPPQ-----GPGQSPMKAEEL 284
QY 65 ---GSLESQPPPIQOTA-CPSSSCLREGEKQNGDSSAGGDFPPPAE---VEPT- 113
DB 285 GEDSAASLSLQSLQGEAQPAPACCEKQESQAEELGPGQGEAEDEDEKVAVASPTPP 344
QY 114 -----EAEELLAQPCHESEASKLGAAPAGGEEEMQOQOLGKK----- 152
DB 345 VSPVNSTPVPAPPEQLSEAA-LKAMEEVAQVLEDDQHNHLLSEKQEQOLREKLCQEE 403
QY 153 -----HRRRPSKKRRMKRYKLTLEEKKKFDEKQSLRASRTAEMFAKQPVAPYNT 205
DB 404 EEEILRLHQOEKSSLSLRELRQKATEEEDARRREESQRLSLRAQVOSSTQA----- 457
QY 206 TQFLMDHDOEERDLKTYGLSKRAAKSD-----DTSDDDFMEEGEGEGGS----- 252
DB 458 -----DEQIADQASLQKLRELEESQOKKARASLEQKNQMLDLQLELEIASGSKSQ 511
QY 253 -----DQMGGDSSEFLQDRFSETYER-----HTE---SLQNMKS 284

[illegible]

```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Fiankooch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doop L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Fertlira S., Fleischmann W.,
RA Foster C., Garfield A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houston D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Schelder F., Shen H.,
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtka R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003435; AAF6083.1; -
DR HSSP: P00730; 2CTC.
DR FlyBase: FBgn0029807; CG3108.
DR InterPro: IPR000834; Zn_cathopept.
DR Pfam: PF00246; Zn_cathopept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
SQ SEQUENCE 1192 AA; 132088 MW; 72408D5C5D3D718E CRC64;

Query Match 7.4%; Score 141.5; DB 5; Length 1192;
Best Local Similarity 22.3%; Pred. No. 0.086;
Matches 80; Conservative 57; Mismatches 138; Indels 83; Gaps 17;
OY 4 PFLSPYQHQPOTSNGTGAAGVDELNPERPPGAEEVPEEDSRWSRAFPQLGGPRGEG 63
DB 252 PLNDELPLPDESPPAATESAV--EELKESAMADQVVESEIOPQVQ---GEYQSES 306
OY 64 EG-----SLESQPPLOTQACPESSCLREGKQNGDSSAGCDFPPAFAVEPTPAE 116
DB 307 DGEQAEKPELEADP---EVEAPEAEQPEAE-----PQLEVEPQPEVE 348
OY 117 LLAQPCHDSEASKIGAPAGEEEMGGQQLGKKHRRRPPYKRLTWEEK 176
DB 349 --SQPEVESQPEVAQPEVEPQSEVSQ-----PEASHSEPE---TQAEVE 390
OY 177 KFDEKSLRASRIAEAMFAKQOPVAPYNTTQFLMDHDHQEERDLKGLTGLYSKRAAKSDT 236

```

```

DB 391 AQPEVESLPEAESQPE--AESQP-----EREPEVE---AEKISDNEYDT 429
OY 237 SDDPFMEGGEGEDSGDGMCGSEFLQDRDSEYTERHYHTESLQMSKQELKEYLEEK 296
DB 430 TEASLMEFLVE--GLEDGLTAMDNLVPEELAESDQKQETE-LESEDDQSPVTEAIE-EQ 485
OY 297 CLSRMEDNNRLRLSKRLGCDGDAVRRELELDRLAENLQ-LITENELHROOERAP 353
DB 486 AVPELEQEKER---EPEDITLAD-----ETREDSQAQPSNPEPVLAEQHTAEIAP 534

RESULT 12
OY 076153 PRELIMINARY; PRT; 443 AA.
AC 076153:
DT 01-NOV-1998 (TREMBLrel. 08. Created)
DT 01-NOV-1998 (TREMBLrel. 08. Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19. Last annotation update)
DE Rsp60.
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattodea; Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193551; PubMed=10727895;
RA Arai T., Kubo T., Natori S.;
RT "Identification, characterization and cDNA cloning of two novel
RT proteins secreted into the external space of the regenerating leg of
RT periplaneta americana."
RL Insect Biochem. Mol. Biol. 30:287-295(2000).
DR EMBL: AB012307; BAA32795.1; -
SQ SEQUENCE 443 AA; 47987 MW; 91AF646B161C99B CRC64;

Query Match 7.4%; Score 140.5; DB 5; Length 443;
Best Local Similarity 20.4%; Pred. No. 0.031;
Matches 80; Conservative 66; Mismatches 140; Indels 107; Gaps 17;
OY 4 PFLSPYQHQPOTSNGTGAAGVDELNPERPPGAEEVPEEDSRWSRAFPQL 55
DB 47 PVAAEVRKREPAAATTTAAADVPKEDKPAEEVPPAPVVEEMKPAEDA----- 96
OY 56 GGRQPEGEGSLESQPPLOTQACPE--SSCLREGKQNGD-----DSSAGCDFPPA 107
DB 97 ---PSPAEAPAEQEVPAV-DEAVPEAVQDETNNKGEVNGQSSDLKTEATSSQDAKPV 152
OY 108 E-----VEPTPAELLQACHSEASKIGAPAGEEEMGGQQLGK 150
DB 153 EEKVALNAPKVGEEPAEKVEEPAEKVEEVVQDQAAPTTAEPKAEEEKPAKDE--K 209
OY 151 KKHRRRP-SKKRRHMKPYKRLTWEEKKKKFKDEKSLRASRIAEAMFAKQOPVAPYNTQFL 209
DB 210 VEBAAPVSRKKRETAP-----KEEEKKPTTKKA-----KAE----- 242
OY 210 MDHHDQEPDLKTLGLYSKRAAKSDDTSDDPFMEGGEGEDGSDGMGDSGSEFLQDRFSE 269
DB 243 -DVAVQEPKQDQVAVQEEVEVQVQEPQDQVQDDAKKE-----EVKEGQAKE 291
OY 270 TYEKYHTESLQMSKQELIKYELLEKCLSMEDNNRLRLSKRLGCDGDAVRRELE 326
DB 292 EEVAVQEEVQKQADAAQVQKQDDV-----KVQDE---VKLEEVKVESDAKEEVKVEAK 341
OY 327 LELDRLAENLQLTLENLIRQ---QERAPLTK 356
DB 342 VEQDAKQVEDKV--QDEVKQDAAAEVNPVSK 372

```

```

RESULT 13
OY 08TXA4 PRELIMINARY; PRT; 609 AA.
AC 08TXA4:
DT 01-JUN-2002 (TREMBLrel. 21. Created)

```

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Uncharacterized protein.
GN MK0771.
OS Methanopyrus kandleri.
OC Archaea: Euryarchaeota; Methanopyrii; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozhavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL: AE010369; AAM01985.1; -
KM Complete proteome.
SQ SEQUENCE 609 AA: 69552 MW: AB10C9780DC5AD78 CRC64;

Query Match 7.3%; Score 140; DB 17; Length 609;
Best Local Similarity 22.0%; Pred. No. 0.049;
Matches 80; Conservative 66; Mismatches 147; Indels 70; Gaps 17;

OY 31 ERPPCAEER-----VPEEDSRMOSRAFPOLGGRPEEGSGS-----LESQPPPLQTQ 77
DB 5 KRKGCKRREPDLKLKPPREGGAPKILKPPREGPG-ESKGAFAVKKLKKP----- 58
OY 78 ACPESSCLREGEKQCGDSSAGDPPEPAVE--PTPAELLAOPCHDSEASKLGAPAAG 136
DB 59 -----KPEKPPSGSEEEKKKKEPPPAELIKPPKEPPRAPSNLAELERLKAENKK 110
OY 137 GEE---EM-----GGOQOLQK-KKRRRPPSKKKRHKMPYKLTWEKKKDE---- 180
DB 111 LREELDEWRNKAASAMGERDLRSEIKRLEKELEKQELDKYIKISKOLEKELKAKRE 170
OY 181 ----KSLRASRIRAFEMFAKGPVAPYNTQFLMDHDQDEPDLKTGLYSKRAAKSDT 236
DB 171 SEELKEKAEYERERKETA-----GKYNELKSKLDLSDONRLAENL--KKLEKYEI 223
OY 237 SDD-DFMEEGEEDGSDGSGDSEFLQROFSEF-YER---YHTESLQNMN--KOEL 287
DB 224 KEERDLKEETKEVGKTK-----DQLAKLQSKLKEVKSERDLANVEALRNENEKLRKI 279
OY 288 IKEYLELEKCLSRMEDENNRLLESKRLGGDARVRELELDRLRAENLQLLTENELHR 347
DB 280 DKLSELSNLOKKLKDREKLEKAROHQIKLREELKRDEETRLKKAOSKL--KDEIKR 337
OY 348 QOE 350
DB 338 YEF 340

RESULT 14
O94DE3 PRELIMINARY: PRT: 714 AA.

AC O94DE3.
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE P0683F02.9 protein.
GN P0683F02.9.
OS Oryza sativa (Rice)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(ca3) genomic DNA, chromosome 1, PAC
RT clone:P0683F02."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003289; BAB63706.1; -
SQ SEQUENCE 714 AA: 77901 MW: 3A99C4DCA9155299 CRC64;

Query Match 7.3%; Score 140; DB 10; Length 714;
Best Local Similarity 21.8%; Pred. No. 0.06;
Matches 74; Conservative 52; Mismatches 132; Indels 82; Gaps 12;

OY 31 ERPPCAEERVPEDSRMOS-RAFPOLGGRPEEGSGSPPPLQTQACPESSCLREGE 89
DB 28 ERADOLAKRVALEBOYQAVTAQRQARATATEVIGLISHQFGNLSVDVDSGDRGE 87
OY 90 KG-----QNGDSSAGCDPPPAEVEPTPEALLAOPCHDSEASKLGAPAAGEEWG 142
DB 88 EDDDPDRARSQDPAVSGSGEOPPAO-----CEAEDALSGTAEPGGGLSW- 133
OY 143 QOOROLGKKHRRRPPSKKKRHKMPYKLTWEKKKFKDEKQSLRASRIRAFEMFAKGPAP 202
DB 134 -KGRVSPRKAROLKQKIRRSY--FYLLSDPSPKYRMGQGRKK-KKEL-SNCKSTAP 188
OY 203 YNTQFLMDHDQDEPDLKTGLYSKRAAKSDTSDDPMEEGEEDG-----SDGM 255
DB 189 -----EEOGRDVEELIAMSQKGOQDSDCTDDGQADMDGFAVGQYVIRYEKDG- 236
OY 256 GGDGSEFLQROFSEFTEYHTESLQNMKQELIKEYLELEKCLSRMEDENNRLLESKRL 315
DB 237 -----EMERYLER-----QAEILQOYEAEAEARQWKEKOPNENSSAK-- 274
OY 316 GGDARVRELELDRLRAENLQLLTENELHROOGRAPLS 355
DB 275 -----VHVEAKKACOITENGMSKHSRLA 300

RESULT 15
O15087 PRELIMINARY: PRT: 1520 AA.

AC O15087.
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE KIAA0383 protein (Fragment).
GN KIAA0383.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN.
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
DR EMBL: AB002381; BAA20837.1; -
DR InterPro: IPR002717; MOZ_SAS.
DR InterPro: IPR001965; ZnF_PHD.
DR Pfam: PF01853; MOZ_SAS; 1.
DR Pfam: PF00628; PHD; 1.
DR SMART: SM00249; PHD; 1.
FT NON_TER 1
SQ SEQUENCE 1520 AA: 171095 MW: B8C666120B0ADEFF2 CRC64;

Query Match 7.3%; Score 139; DB 4; Length 1520;
Best Local Similarity 18.6%; Pred. No. 0.17; Indels 150; Gaps 16;
Matches 84; Conservative 67; Mismatches 150; Indels 150; Gaps 16;

OY 7 SEYOHQOTSNGTG-----AAVQELNPERPGAEERVPEDSRMOSRAFPOLG 56

QY 41 ProGluGluAspSerArgrTrpGlnSerArgrAlaPheProGluLeuGlyGlyArgrProGly 60
|||||
Db 246 CCCGAGGAGACAGTACGTGGCAATCGACAGCGCTTCCCGCAGTTGGGTGGCGCTCCGGG 305
QY 61 ProGluGlyGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
|||||
Db 306 CCGGAGGGGGAAGGAGCGCTGGAATCCCAACCCACTCCCTTGCAGACCCAGCGCTGCCA 365
QY 81 GluSerSerCysLeuArgrGluGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
|||||
Db 366 GAATCTAGCTGCTGTGAGAGGGGCGAAGGGCCAGAAATGGAGACACTGCTCCGCTGGC 425
QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuAlaGln 120
|||||
Db 426 GGGGACTTCCCGCGCGCGGAGAGTGGAAACCCAGCCCGAGCGCTCGCCGAG 485
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGluGlu 140
|||||
Db 486 CCTTGTATGACTCCGAGGCGCAAGTTGGGGCTCTCCGCGAGG-GCGAAGAGGAG 544
QY 141 TrpGlyGlnGlnGlnArgrGlnLeuGlyLysLysLysHisArgrArgrProSerLysLys 160
|||||
Db 545 TGGGACAGCAGCAGACAGAGCTGGGAAAGAAACATAGAGAGCGCCGCTCCAAAG 604
QY 161 LysArgHisTrpLysProTrpTrpLysLeuThrTrpGluGluLysLysPheAspGly 180
|||||
Db 605 AAGCGGATTTGGAAACCGTACTCAACGCTACCTCGGAGAGAGAAAGAAAGTTCCGACGAG 664
QY 181 LysGlnSerLeuArgrAlaSerArgrAlaArgAlaGluMetPheAlaLysGlyGlnProVal 200
|||||
Db 665 AAACACAGACCTTCGAGCTTCAAGATCCGAGCGAGATGTTCCCAAGGCGCCAGCGGCTC 724
QY 201 AlaProTrpArgrsnhrTrpGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
|||||
Db 725 GCCCCCTATACACCCAGCGGCTTCATGATGATGATCAGACAGGAGGAGCGCGAGATCTC 784
QY 221 LysThrGlyLeuTrpSerLysArgrAlaAlaAlaLysSerAspAspThrSerAspAsp 240
|||||
Db 785 AAACCGCGCTGTACTCTCAAGCGCGCGCGCCCAATCCGACGACACACGATGACGAC 844
QY 241 PheMetGluGluGlyGlyGluGluAspGly 250
|||||
Db 845 TTTCATGGAARACGGCGTGGAGAGATGGC 874

RESULT 3
Bi871190 771 bp mRNA linear EST 11-OCT-2001
LOCUS 603395081F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5404724 5',
DEFINITION mRNA sequence.
Bi871190
ACCESSION Bi871190.1 GI:16044865
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 771)
NIH-MGC htlp://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM12032 row: b column: 21
High quality sequence stop: 733.
Location/Qualifiers

source 1. .771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5404724"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="Dh10b (phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."

BASE COUNT 187 a 213 c 273 g 98 t

ORIGIN

Alignment Scores:
Pred. No.: 1.61e-87 Length: 771
Score: 1266.50 Matches: 248
Percent Similarity: 96.51% Conservative: 1
Best Local Similarity: 96.12% Mismatches: 5
Query Match: 66.31% Indels: 5
DB: 13 Gaps: 1

US-09-972-758a-2 (1-359) x Bi871190 (1-771)

QY 21 AlaAlaAlaValGlnGlnGluLeuAsnProGluArgrProProGlyAlaGluArgrVal 40
|||||
Db 3 GCTGCTGCTCTCCAGAGAGAGCTGAACCTCGAGCGCCCGCGGAGGAGCGGGT 62
QY 41 ProGluGluAspSerArgrTrpGlnSerArgrAlaPheProGluLeuGlyGlyArgrProGly 60
|||||
Db 63 CCGGAGGAGGACAGTACGTGGCAATCGACAGCGCTTCCCGCAGTTGGGTGGCGCTCCGGG 122
QY 61 ProGluGlyGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
|||||
Db 123 CCGGAGGGGGAAGGAGCGCTGGAATCCCAACCACTCCCTTGAGAGCCAGCGCTGTGCCA 182
QY 81 GluSerSerCysLeuArgrGluGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
|||||
Db 183 GAATCTAGCTGCTCGAGAGAGGGCGAGAAAGGCCACAGATGGAGCACACTGCTCGCTGGC 242
QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
|||||
Db 243 GCGGACTTCCCGCGCGCGGAGAGTGGAAACCCAGCCCGGAGCGGAGCTTCGCCAG 302
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGln 140
|||||
Db 303 CCTTGTATGACATCCAGCGGCGAGTGGGGCTTCCTCCCGAGGGCGGCAAGAGGAG 362
QY 141 TrpGlyGlnGlnGlnArgrGlnLeuGlyLysLysLysHisArgrArgrProSerLysLys 160
|||||
Db 363 TGGGACAGCAGCAGACAGACGCTGGGAGAGAAAGAAATAGGAGACCCCGCTCCAAAGAG 422
QY 161 LysArgHisTrpLysProTrpTrpLysLeuThrTrpGluGluLysLysLysPheAspGly 180
|||||
Db 423 AAGCGGATTTGGAACCGTACTCAAGCTGACCTCGGGAACAGAAAGAAAGTTCCAGAG 482
QY 181 LysGlnSerLeuArgrAlaSerArgrAlaArgAlaGluMetPheAlaLysGlyGlnProVal 200
|||||
Db 483 AAACAGAGCGCTTCGAGCTTCAAGATCCGAGCGGAGATGTTCCCAAGGGCGGCGGCTC 542
QY 201 AlaProTrpArgrsnhrTrpGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
|||||
Db 543 GCGCCCTATACACCCAGCGGCTTCATGATGATGATCAGCAGCAGAGGAGCGGATCTC 602
QY 221 LysThrGlyLeuTrpSerLysArgrAlaAlaAlaLysSerAspAspThrSerAspAsp 240
|||||
Db 603 AAACCGCGCTGTACTCAAGCGGGCGC-GCCAAATCCGACGACACCACTGATGACGAG 661
QY 241 PheMetGluGluGlyGlyGlyGluGluAspGlyGlyLysSerAspGlyMetGly---GlyAspGly 259
|||||
Db 662 TTTCATGGAAGAGGGGTTGAGAGGATGGGCGGCGGAGCGGATGGAGAGCGGCGGCT 721

OY	260	SerGUglnArgasp-PheSerGUgln-TTrTYrGUglnACTyr	274
Db	722	AGCGAGGTTCTGCCAGCGGGAACCTTCTGGAGAACGTACGAGCCGTAC	769
RESULT 4			
LOCUS	BF984049		
DEFINITION	602306989P1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4398238 5',		
ACCESSION	BF984049		
VERSION	BF984049.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukariota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia, Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strusberg, Ph.D.		
	Email: cgapbs@mai.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: LNL0100 row: a column: 23		
	High quality sequence stop: 691.		
FEATURES			
source	Location/Qualifiers		
	1..982		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4398238"		
	/clone_lib="NIH_MGC_88"		
	/tissue_type="duodenal adenocarcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: small intestine; Vector: pCMV-SPORT6;		
	Site_1: NotI; Site_2: SalI; Cloned unidirectionally;		
	oligo-dn primed. Average insert size 1.767 kb. Library		
	enriched for full-length clones and constructed by Life		
	Technologies. Note: this is a NIH-MGC Library."		
BASE COUNT	261 a 235 c 320 g 166 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,05e-84	Length:	982
Score:	1231.00	Matches:	238
Percent Similarity:	99.17%	Conservative:	0
Best Local Similarity:	99.17%	Mismatches:	2
Query Match:	64.45%	Indels:	1
DB:	12	Gaps:	0
US-09-972-758A-2 (1-359) x BF984049 (1-982)			
OY	120	GlnProCYSHisASPserGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlnGln	139
Db	8	CAGCCTTGTCATGACTCCGAGCCAGTGAAGTTGGGGCTCCTGCCGCGAGGGGCGAAGAG	67
OY	140	GluTPGGLyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLys	159
Db	68	GAGTGGGGACACGACACAGAGAGCTGGGGAGAAAAACAATAGAGACGCCGTCCAG	127
OY	160	LysLysArgHisTPPLysProTYrTYrLysLeuThrTPGGLyGlnLysLysPheASP	179
Db	128	AAGAAGCGCGACTTGGAAACCGTACTACAGCTGACTCGGAAGAGAAACAAAGTTGCAC	187
OY	180	GluLysGlnSerLeuArgAlaSerArgTlleArgAlaGlnMetPheAlaLysGlyGlnPro	199
Db	188	GAGAAACGAGACCTTCGAGCTTCAAGAGATCCGAGCCGAGATGTTCCGCAAGGCGACCGC	247

QY	200	Val1AlapPcYrAspTThrThG1nPhemeuMeTsAspSPH1sAspG1nG1u1PrAsp	219
Db	248	GTCCGCGCCCTATTACACCCAGCACTTCTCTGTGATATATCAGACACGAGAGACCCGAT	307
QY	220	Leu1sYThrG1yLeuYrSerYserysArg1a1a1a1a1ySerAspSPThrSerAspSP	239
Db	308	CTCAAAACCCGCCCTGTACTCTCCACCGCGCCGCCCAATCTCCAGCAGCACCCGATATAC	367
QY	240	AspPheMeTg1uG1nG1yG1yG1nG1uAspD1yG1ySerAspG1yMeTg1yG1yAspG1y	259
Db	368	GACTTCATGAGAGAGGGGCTGACAGAGATGGGGCACCGATGGGATGGAGGGGACGGC	427
QY	260	SerG1uPheLeuG1n1aYAspSPHeserG1uThrYrG1uAryG1yTh1sThG1uSerLeu	279
Db	428	AGCGAGTTTTCGACGGGCACTTCTCGGAGACGTACGAGCGGCTACCAACAGCAGACGCTG	487
QY	280	G1nAspMeTserYrG1nG1uLeu1le1ySg1uYrLeuG1uLeuG1u1yScyLeuSer	299
Db	488	CAGAACATAGACCAAGCAGACGACTCATCAAGAGTACCTCGGAACTCGGAATGCTCTCG	547
QY	300	ArgMeTg1uAspG1uAsnAsnArgLeuArgLeuG1uSerYserysArgLeuG1y1AspAsp	319
Db	548	CGCATGAGAGACGAGAACACCCGCTGCGGCTGGAGACACAGCGGCTGGTGGCGGACGAC	607
QY	320	AlaArgValArgG1uLeuG1uLeuG1uLeuAspArgLeuArg1aG1aG1uAsnLeuG1u	339
Db	608	GCGGGTGTGCGGAGCTGAGCTGTGAGCTGTGACCGGCTGCGCGCGGAGAACCTCCACTG	667
QY	340	LeuThrG1uAsnG1uLeuH1sArgG1nG1nG1uArg1a1aProlSerYsPheG1yAsp	359
Db	668	CTGACCGAGAACGAACTCCACCGGACGAGAGGAGCGGCG-CTTCCCAAGTTTGGAGAC	726
RESULT 5			
LOCUS	BC819109	797 bp	mRNA
DEFINITION	602781202p1 NCI_CGAP_Brn67 Homo sapiens cdna clone IMAGE:4932025	linear	EST 22-MAY-2001
ACCESSION	BC819109		
VERSION	BC819109.1	GI:14166696	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 797)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: scapbs-rt@mail.nih.gov Tissue Procurement: David N. Louis, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: LLAM10857 row: C column: 02 High quality sequence stop: 797.		
FEATURES			
source			
	1..797		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4932025"		
	/clone_lib="NCI_CGAP_Brn67"		
	/tissue_type="anaplastic oligodendroglioma with 1p/19q loss"		
	/lab_host="DH10B (T1 phage-resistant)"		
	/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; Note: Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: a NCI_CGAP Library."		
BASE COUNT	207 a 223 c 248 g 119 t		

QY 261 GluPheLeuGlnAArgspPheSerGluThrTyrGluAArgTyrHisThrGlnSerLeuGln 280
|||||
Db 428 GAGTTTCTGCGAGCGACTTCTCGAGAGCTAGAGCGGTACACAGGAGACCTCGACG 487
QY 281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGlnLeuGlnLysCysLeuSerArg 300
|||||
Db 488 AACATAGCAGAGGAGGACATCAAGAGTACTGACCTGACGAGAAAGTCCCTTCGCGC 547
QY 301 MetGlnAspGlnAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnLysAspAspAla 320
|||||
Db 548 ATGAGAGACAGACAGAACACCGCGCTGCNCGTGACAGACGCGCTGGTGGCGACGACCG 607
QY 321 ArgValArgGluLeuGlnLeuGlnLeuAspArgLeuArgAlaGlnLysLeuGlnLeu 340
|||||
Db 608 CCTGTGCGGGAGCTTGAGCTTGACCTTGACCTTGNNCGCCGACACCTTCACACTGCTG 667
QY 341 ThrGlnAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
|||
Db 668 ACCGANAACGACCTGACCCGACAGCAGAGCGAGCGCGCTTCCAAAG 715
RESULT 7
BI091005 852 bp mRNA linear EST 20-JUN-2001
LOCUS 602853667F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4995065 5',
DEFINITION mRNA sequence.
ACCESSION BI091005
VERSION BI091005.1 GI:14509335
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 852)
NIH-MGC http://mgc.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM1018 row: e column: 18
High quality sequence stop: 700.
FEATURES
source location/Qualifiers
1..852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4995065"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: PCMV-SPORT6; Site:1: NOTI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT 238 a 196 c 286 g 132 t
ORIGIN
Alignment Scores:
Pred. No.: 2..88e-76 Length: 852
Score: 1120.00 Matches: 226
Percent Similarity: 98.70% Conservative: 1
Best Local Similarity: 98.26% Mismatches: 3
Query Match: 58.64% Indels: 3
DB: 13 Gaps: 0
US-09-972-758A-2 (1-359) x BI091005 (1-852)
QY 130 LeuGlnAlaProAlaIaGlyGluGluGluTPrGlyGlnGlnArgGlnLeuGly 149

Db 5 GTGGGGGCTCTCCCGCAGCGGCGCAAGAGAGTGGGAGCAGACAGACAGCTGGCG 64
QY 150 LysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrLys 169
|||||
Db 65 AAGAAAAACATVAGAGACCCCGCTCAAGAAAGAACCGGCA-TGGAAACCGTACTVCAAG 123
QY 170 LeuThrTrpGluGluLysLysLysPheAspGlnLysGlnSerLeuArgAlaSerArgIle 189
|||||
Db 124 CTGACCTGGGAGACAGAAAGAAAGCTTCAGACAGAAACAGACGCTTGACCTCAAGATC 183
QY 190 ArgAlaGluMetPheAlaLysGlnProValAlaProTyrAsnThrThrGlnPheLeu 209
|||||
Db 184 CGAGCCGAGATGTCTCCCAAGGGGCCAGCGGTGCCCCCTATPACACACAGCGAGTTCCTC 243
QY 210 MetAspAspHisAspGlnGlnGluProAspLeuLysThrGlyLeuLysSerLysArgAla 229
|||||
Db 244 ATGATGATATCACCACACAGACGAGCGGATCTCAAAACCGGCTTGACTCTCAAGCGGGCC 303
QY 230 AlaAlaLysSerAspAspThrSerAspAspPheMetGluGluGlyGluGluAsp 249
|||||
Db 304 GCCGCCAAATCCGACGACACACAGCATGACACTTCATGGAAGAGAGGGGTAGAGGAT 363
QY 250 GlyGlySerAspGlyMetGlyLysAspGlySerGluPheLeuGlnAArgspPheSerGlu 269
|||||
Db 364 GGGGGCAGCGATGATGATGGAGGGGACGCGACGAGCTTCG-CAGCGGAGCTTCTCGAG 422
QY 270 ThrTyrGluArgTyrHisTrpGlnSerLeuGlnAsnMetSerLysGlnGluLeuIleLys 289
|||||
Db 423 ACCTACGAGCGGTACACACAGGAGACCTGCACAAATGACAGAGAGGAGCTCATCAAG 482
QY 290 GluTyrLeuGlnLeuGlnLysCysLeuSerArgMetGlnAspGlnAsnAsnArgLeuArg 309
|||||
Db 483 GATTACTCTGGAAGTGGAGAAAGTCCCTCGCGCATGAGAGACAGAAACCGCTCGCG 542
QY 310 LeuGlnSerLysArgLeuGlnGlyLysAspAspAlaArgValArgGluLeuGlnLeu 329
|||||
Db 543 CTGGAGAGCAGAGCGGCTGGTGCGACGACGCGCTGTGGGAGCTGGAGCTGGAGCTG 602
QY 330 AspArgLeuArgAlaGlnAsnLeuGlnLeuLeuThrGlnAsnLysLeuHisArgGlnGln 349
|||||
Db 603 GACCGGCTGGCGGCCAGAACTCCAGCTGCTGACGACGACGACGACGACGACGACG 661
QY 350 GluArgAlaProLeuSerLysPheGlyAsp 359
|||||
Db 662 GACGAGAGCGCGCTTCCAAAGTTTGGAGAC 691
RESULT 8
BG291150 871 bp mRNA linear EST 21-FEB-2001
LOCUS 602387182F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4515812 5',
DEFINITION mRNA sequence.
ACCESSION BG291150
VERSION BG291150.1 GI:13048811
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 871)
NIH-MGC http://mgc.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM10406 row: d column: 21

Db 123 CCACTCCCTTGACAGCCAGCCCTGTCAGAAATCTAGCTGCTGAGAGAGGCGGAGAG 182
 QY 91 GYGLInasnglyAspAspSerSerAlaGlyIAspPheProProAlaGluValGlu 110
 Db 183 GGGCAGATGGGAGCAGCTGCTCGCTGGGGGAGACTCCCGCGCGGAGAGAGTGGAA 242
 QY 111 ProThrProGluAlaGluLeuAlaGluProCysHisAspSerGluAlaSerLysLeu 130
 Db 243 CCGACGCCCGGAGCGCGCTGCTGCGCCAGCCTTGTCATGACTCCGAGCCAGTAAGTTG 302
 QY 131 GYAlaProAlaAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 150
 Db 303 GGGGCTCTGCTCCCGAGGGGCGAAGAGAGTGGGAGACAGACAGACAGCTGGGGAG 362
 QY 151 LysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeu 170
 Db 363 AAAAACATGAGAGAGCGCCGCTCCAGAAAGACGGCATTTGGAACCGTACTCAACCTG 422
 QY 171 ThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190
 Db 423 ACCTGGGAAGAGAAAGAAAGTTGAGAGAAACAGAGCCTTGAGCTTCAAGCATCCGA 482
 QY 191 AlaGluMetPheAlaLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 210
 Db 483 GCGAGATGTTCCCAAGGGCGACGCGGTCGCGCCCTATACACACAGCTTCTCAT 542
 QY 210 TAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArg-Ala 230
 Db 543 GGAATGATCAGACAGAGAGAGCGGATCTCCNACCGGCTGACTCCAGCGGGGCGG 602
 QY 230 la-AlaLysSerAspAspThrSerAspAsp-----AspPheMetGluGlyGly 246
 Db 603 CCGCCCAATCCCGGAGACACCGACGCGATTTGACCATTCATGGAANCAAGGCGGTG 662
 QY 247 GlnGluAspGlyGly---SerAspGlyMetGlyGlyAspGlySerGlu----- 261
 Db 663 AAGAAAGATGGGGGCGCGGATGGGAATGGGGAAGGGGAACGCGCGCCCAATTTT 722
 QY 262 PheLeuGln 264
 Db 723 TTCTTCAG 731
 RESULT 10
 AL578242/c 1001 bp mRNA linear EST 16-FEB-2001
 LOCUS AL578242 LTI_NFL006_P12 Homo sapiens cDNA clone CS0DK002YF07 3
 DEFINITION prime, mRNA sequence.
 ACCESSION AL578242
 VERSION AL578242.1 GI:12942132
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1001)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..1001
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DK002YF07"
 /clone_lib="LTI_NFL006_P12"
 /issue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com

BASE COUNT 218 a 289 c 218 g 271 t 5 others
 ORIGIN

Alignment Scores:
 Pred. No.: 3 8e-72 Length: 1001
 Score: 1067.00 Matches: 209
 Percent Similarity: 97.66% Conservative: 0
 Best Local Similarity: 97.66% Mismatches: 5
 Query Match: 55 86% Indels: 1
 DB: 9 Caps: 0

US-09-972-758a-2 (1-359) x AL578242 (1-1001)

QY 146 ArgGlnLeuGluLysLysLysHisArgArgArgProSerLysLysArgHisTrpLys 165
 Db 999 AGACAGCTGGGAGAAAGAAACATAGAGAGCGCCGTCAGAAAGAGCGGCTTGGAAA 940
 QY 166 ProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArg 185
 Db 939 CCTACTACAACTGACCTGGAGAGAGAAAGTTCAGCAGAAACAGACCTTCCGA 880
 QY 186 AlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThr 205
 Db 879 GCTTCAAGATCCGACCGGAGATGTTCCCAAGGGCGAGCGGCTCCCTTAACACC 820
 QY 206 ThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyr 225
 Db 819 AGCAGTCTCTCATGATGATGATCATCAGACACGAGAGACCGGATTTCAAAACCGGCTGTAC 760
 QY 226 SerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGly 245
 Db 759 TCCAGACGGGCGCCCGCCCAATCCGACGACACCGAGCATGACATTCATGGAAGAGG 700
 QY 246 GlnGluGluAspGlyLysSerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArg 265
 Db 699 GGTGAGGAGGATGGGGGACACCGATGGGATGGGAGGACGAGGATTTCTGACAGCGG 640
 QY 266 AspPheSerGlnThrTyrGluArgTyrHisThrGlnSerLeuGlnAsnMetSerLysGln 285
 Db 639 GACTTCTCGAGACGATGACGAGCGGTACACACGAGAGCGCTCGAGAAATGACCAAGCAG 580
 QY 286 GlnLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsn 305
 Db 579 GAGCTCATCAAGAGATCACTGGAACCTGGAAGAGTGCYCTCCGGGCTGGAGGACGAGAAC 520
 QY 306 AsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeu 325
 Db 519 AACCGGCTCGGCTGGAGACAAAGCGGTGGGGGAGCTACCGCGCTGTGCGGAGCTG 460
 QY 326 GlnLeuGlnLeuAspArgLeuArgAlaGluAsnLeuGlnLeuThrGlnGlnGluLeu 345
 Db 459 GAGCTTGAGAGCTGAGCGCGTG-CGCCCGGAGAACTTCAGGTGTGACCGAGAAACGAC 401
 QY 346 HisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
 Db 400 CACCGGACAGAGAGAGCGAGCGCGCTTCCAAAGTTGGAGAC 359
 RESULT 11
 BM474949 1080 bp mRNA linear EST 05-FEB-2002
 LOCUS BM474949
 DEFINITION AGENCOURT.6476734 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562756
 5', mRNA sequence.
 ACCESSION BM474949
 VERSION BM474949.1 GI:18523991
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1080)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLMI2292 row: k column: 13
High quality sequence stop: 404.
Location/Qualifiers
1..1080
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5562756"
/clone_lib="NIH-MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dt primed. Average insert size 1,767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH-MGC Library."
BASE COUNT 251 a 339 c 357 g 133 t
ORIGIN
Alignment Scores:
Pred. No.: 2,59e-71 Length: 1080
Score: 1056.50 Matches: 222
Percent Similarity: 88.21% Conservative: 10
Best Local Similarity: 84.41% Mismatches: 20
Query Match: 55.31% Indels: 13
DB: 13 Gaps: 4
US-09-972-758a-2 (1-359) x BM474949 (1-1080)
QY 31 GIUATGPROGGLYALAGLUAGLUARVALPROGLUGLUASPSERARGTGINSERAR 50
|||||
Db 3 GAGCGCGCCCGCAGCGCGGAGGAGCGCGTCCCGAGAGACAGTAGTGCGCATGACA 62
QY 51 ALAPHEPROGGLUGLYLARGPROGLYPROGLUGLYGLYSERLEUGSERGIN 70
|||||
Db 63 GCGTTCCCGCAAGTTGGGTGGCGTCCGGCGCGGAGGCGGAGGCGAGCTCGAATCCAA 122
QY 71 PROPRPROLEUGLTHRGILALACYSPROGLUSERSERCYSLAURGLUGLYGULYS 90
|||||
Db 123 CCACCTCCCTTCAGACGCCAGCGCTCTCCAGATCTAGCTGCTTGAGAGGCGCGGAG 182
QY 91 GLYLINASNGLYASPAASPSERSERALAGLYGLYASPPHEPROPROALAGLUVALGU 110
|||||
Db 183 GCGCAGAAATGGGAGCAGACTGTCGCGTGGCGGAGACTTCCGCGCGCGCAGAGAAGTGC 242
QY 111 PROTHRPROGLUALAGLULEULAGLNPROCYSYHASPERSERGLUALASERLYSLEU 130
|||||
Db 243 CCGACGCGCGGAGCGAGCTGCTCCGCCAGCTTGTCATGACTCCAGGCCAGTAGTGG 302
QY 131 GYALAPROALALAGLYLUGLUGLUGLUTPRGLYGLNGLNARGLNLNUGLYLS 150
|||||
Db 303 GGGGCTCTCGCGCAGGGGGCGAAGAGAGTGGGACACACAGCAGAGACAGCGGGGAG 362
QY 151 LYSLSYHSATGARGARGPROSERLYSLYSARHISTRPLYSPTGYTYRFLYSLEU 170
|||||
Db 363 AAAAAACATAGGAGAGCGCGCTCCAAAGAGAGCGGCAATTGGAAACCGTACTACAGCTG 422
QY 171 THRTTPGLUGLUYLSYLSYPHEASPGLUYSGINSERLAURGLASERATGLIARG 190
|||||

Db 423 ACCTGGAGAGAGAGAA-AAGTTGACGAGAAACAGGCTTCGAGCTTCAAGATCCGA 481
QY 191 ALAGLMEtPHEALALYSGLYGLNPROVALIAPROTYRASNThrClnPheLemmet 210
|||||
Db 482 GCCGAGATTTTCCCGCAGGGCCAGCCGTCGGCCCTATAC-ACCAGCAGTTCTCATG 540
QY 211 ASPASPHISASPOLNGLUGLUPROASPLEULYSTRH-GLYLEUTYRSErLYSARG-ALIA 230
|||||
Db 541 GATGATCACGACCCAGAGAGACCGGATCTCAAAACCGGCGCTGTACTCCACGGGGCGG 600
QY 230 LALALYSERISASPTHRSERASP-ASPAAP-PHEMET---GLUGLUGLYGLUGL 248
|||||
Db 601 CCGCCAAATCCACGACGACACCGAATGACGACCTTCATGCGGAAAAAAGGCTGGAAGCA 660
QY 248 VASPGLYGLYSERASP---GLYMEtGLYVASPGLYSERGLU---PHEAUGClnARGAS 266
|||||
Db 661 GGATGGGGGCCACCCCATGGGAGATGGAGAGGGAGCGCAACCAAGTTTCTTGCGAGGGCA 720
QY 266 PHESERGLUTHrTYRGLUARGTYRHISTHR-----GLUSERLEUGLNASME 282
|||||
Db 721 CCTTTCGCCGAGAGACGACCAAGGCGGACCCACCGCGGAAAAAGACCGCGCAAAACCT 780
QY 282 t 282
Db 781 G 781
RESULT 12
BG165450 949 bp mRNA line EST 06-FEB-2001
LOCUS 602345830F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4456062 5'
DEFINITION
BGI65450
BGI65450
ACCESSION
BGI65450
VERSION
BGI65450.1 GI:12672153
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 949)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLMI0250 row: k column: 07
High quality sequence stop: 622.
Location/Qualifiers
1..949
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4456062"
/clone_lib="NIH-MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT 243 a 257 c 335 g 114 t
ORIGIN
Alignment Scores:
Pred. No.: 2,96e-70 Length: 949
Score: 1042.00 Matches: 239
Percent Similarity: 89.49% Conservative: 8

Best Local Similarity: 86.59% Mismatches: 21
 Query Match: 54.55% Indels: 13
 DB: 12 Gaps: 1

US-09-972-758a-2 (1-359) x BG165450 (1-949)

QY 77 GlnAlaCysProGluSerSerCysLeuArgGluGluLysGluGlnAsnGlyAspAsp 96
 |||
 Db 3 CAGGCTGTCACAAATCTACCTGCTGAGAGAGCGCAGAGGCCAGAAATGGGAGACG 62
 |||
 QY 97 SerSerAlaGlyGlyAspPheProProAlaGluValGluProThrProGluAlaGlu 116
 |||
 Db 63 TCCTCCGCTGGCGCGCACTCCCGCGCGCGCAGAAATGGAAACGAGCGCCGAGGC- GAG 121
 |||
 QY 117 LeuLeuAlaGluProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGly 136
 |||
 Db 122 CTGCTGCCCCAGGCTTGCTGATGACTCCGAGGCCAGTAAGTG- GGGCTCTCTGCGCAGGG 180
 |||
 QY 137 GlyGluGluGluTrpGlyGlnGlnArgGluLeuGlyLysLysLysHisArgArgArg 156
 |||
 Db 181 GCGGAAGAGAGTGGGGACAGCAGACAGACAGCTGGGGAGAAAAACATAGAGAGCGC 240
 |||
 QY 157 ProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLys 176
 |||
 Db 241 CCCTCCAGAGAAAGAGCGCA- TGGAAACCTTACTACAAAGCTGAGCTGGGAGAGAGAAA 299
 |||
 QY 177 LysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLys 196
 |||
 Db 300 AAGTTTCGACAGAAACAGAGACCTTCGAGCTTCAAGATCCGAGCCGAGATGTTCCGCCAG 359
 |||
 QY 197 GlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetLysAspHisAspGlnGlu 216
 |||
 Db 360 GGCACACCGGCTCGCGCCTATATACACACGACAGCTTCCTCATGATGATCAGCAGCAGAG 419
 |||
 QY 217 GluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThr 236
 |||
 Db 420 GAGCCGGAATCTCAAAACCGGCTGTACTCCAAACGGGCGG- CGCAAAATCCGACGACACC 477
 |||
 QY 237 SerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyLysSer-AspGlyMetGlu 256
 |||
 Db 478 AGGATGACGACACTTCATGAGAGAGGGGGTGAAGAGAGCTGGGGCAGCCGATGGATGGG 537
 |||
 QY 256 YglIAspGlySerGluPheLeuGlnArgAspPheSer- GluThrTyrGluArgTyrHis- 275
 |||
 Db 538 AGGGGACGAGAGC- GAGTTTCTGACAGGGGAGCTTCTGGCAGAGCTGAGACCGGTACACAC 596
 |||
 QY 276 ThrGluSerLeuGlnAsnMetSerLys---GlnGlu-LeuIleLysGluTyrLeuGluLe 294
 |||
 Db 597 ACGGAGAGCCTGCAAGAACATGAGCCCAAGCCGGGAAGCTCATCCAAAGGACGTACTGGACC 656
 |||
 QY 294 uGluLysCysLeuSerArgMetGluAsp-GluAsn-AsnArgLeuArgLeuGlnSerLys 313
 |||
 Db 657 TGAAGAGTTGCTCTGCGCATGAGAGACCGAGAACCAACCGGCTGAGGCTGGAGACCAAG 716
 |||
 QY 314 ArgLeuGlyGlyAspAspAla-ArgValArgGluLeuGluLeuGluLeuAspArgLeuArg 333
 |||
 Db 717 CGGCTGGTGGCGACACGCGCGCTGTGGGAGCTGAGAGCTGGAAGCGGAGCCGCTTGG 776
 |||
 QY 333 gAlaGluAsnLeuGlnLeuLeuThrGluAsnGlu 344
 |||
 Db 777 GCGCGGAGACCTTCAGTGTGAGCGAGAACCA 810
 |||

RESULT 13
 AL576112/c 986 bp mRNA linear EST 16-FEB-2001
 LOCUS AL576112 LRL_NFL006_PL2 Homo sapiens cDNA clone CSDD1072YL06 3
 DEFINITION prime, mRNA sequence.

ACCESSION AL576112
 VERSION AL576112.1 GI:12937932
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 986)
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES
 source
 1. 986
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSDD1072YL06"
 /clone_1lb="LRL_NFL006_PL2"
 /tissue_type="Placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 216 a 284 c 217 g 265 t 4 others
 ORIGIN

Alignment Scores:

Pred. No.: 2,93e-66 Length: 986
 Score: 990.00 Matches: 197
 Percent Similarity: 95.26% Conservative: 4
 Best Local Similarity: 93.36% Mismatches: 8
 Query Match: 51.83% Indels: 3
 DB: 9 Gaps: 1

US-09-972-758a-2 (1-359) x AL576112 (1-986)

QY 149 GlyLysLysLysHisArgArgProSerLysLysLysArgHisTrpLysProTyrTyr 168
 |||
 Db 985 GGAAGAAACATRGAAACCGCGTCCA-----AGAGAAGCGGCAATTGAACCGTACTAC 932
 |||
 QY 169 LysLeuThrTrpGluGluLysLysLysPheAspGlyLysGlnSerLeuArgAlaSerArg 188
 |||
 Db 931 AAGCTGACCTGGGAAAGAAAGAAAGTTCTGACGAAACAGAGCCTTGAGCTTCAAG 872
 |||
 QY 189 IleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPhe 208
 |||
 Db 871 ATCCGAGCGAGATGTTCCGCCCAAGGGCCAGCGGTCGCCCTCATTAACACACGACGATTC 812
 |||
 QY 209 LeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArg 228
 |||
 Db 811 CTCATGATGATATACCAACAGAGAGCGGATCTCAAAACCGGCTGTACTTCAAGCGG 752
 |||
 QY 229 AlaAlaAlaLysSerAspAspThrSerAspAspPheMetGluGluGlyGlyGluGlu 248
 |||
 Db 751 GCGCGCGCAAAATCCAGCAGCACCAGCATGAGACTTTCATGGAAGAGGGGTGAGAG 692
 |||
 QY 249 AspGlyGlySerAspGlyMetGlyLysAspGlySerGluPheLeuGlnArgAspPheSer 268
 |||
 Db 691 GATGGGGCAGCATGATGAGTGGAGGCGCAGCGACGACGAGTTCCTGACGGGACTTCTCG 632
 |||
 QY 269 GluThrTyrGluArgTyrHisTrpGluSerLeuGlnAsnMetSerLysGlnGluLeuIle 288
 |||
 Db 631 GAGACTTACAGCGGATACCAACAGAGACCTTCACAAATGAGCAAGAGAGGCTCAATC 572
 |||
 QY 289 LysGluTyrLeuGluGluGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeu 308
 |||
 Db 571 AASGATTAAGTGAAGTGAAGAGTCTCTCGCGCATGAGAGAGCAGAAACCGGCTG 512
 |||
 QY 309 ArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluGluGlu 328
 |||

Db		511	CGCGTGGAGACCAAGCGGCTGTGGTCGCAGACACGCCGCTGTGCGGAGACTGGAGCTGGAGC	452
Oy		329	LeuAsPaRgLeuArGaIuAAsnLeuGlnleuleuthrGiLuanGluLeuHISArgGln	348
Db		451	CTGACCGCGCTG-CGCGCGCAGAACCTCCAGCTGCTGACCGAGAACGAACTCACCGGACG	393
Oy		349	GInGUArGaIaProlSerLySPheGlyasp	359
Db		392	CAGAGCGCAGCGCCGCTTTCCAACTTGTCAGAC	360
RESULT 14				
BMI29736				
LOCUS	BMI29736	557 bp	mRNA	linear EST 12-MAR-2002
DEFINITION	lf22e01.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens cDNA clone IMAGE:5677152 5' similar to TR:094992 094992 HIS1 PROTEIN. ; mRNA sequence.			
ACCESSION	BMI29736			
VERSION	BMI29736.1	GI:17124288		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Keastner,K., Lemishka,I., Scaerac,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A., Schmitt,A., Theising,B., Rittler,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McCann,R., Cole,R., Tsagarelisvilli,R., Williams,T., Jackson,Y. and Bowers,Y.			
AUTHORS	Endocrine Pancreas Consortium Unpublished (2000) Other-ESTs: lf22e01.x1			
TITLE	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue			
JOURNAL	Endocrine Pancreas Consortium			
COMMENT	Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138 Tel: 617-495-1812 Fax: 617-495-8557 Email: dmelton@biohp.harvard.edu Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: Juliana Brown (brownjef@s.harvard.edu) This sequence now available from the IMAGE consortium, for clone orders contact: infoimage.llnl.gov High quality sequence stop: 443. Location/Qualifiers 1..557 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5677152" /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1" /sex="Both /tissue_type="Islets of Langerhans" /dev_stage="Adult" /lab_host="DH10B" /note="Organ: Pancreas; Vector: pSPORT1; Site1: Not 1; Site_2: Sat 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT 05 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."			
BASE COUNT	142 a	150 c	193 g	72 t
ORIGIN				

Alignment Scores:			
Pred. No.:	2.02e-64	Length:	557
Score:	963.00	Matches:	184
Percent Similarity:	99.46%	Conservative:	0
Best Local Similarity:	99.46%	Mismatches:	1
Query Match:	50.42%	Indels:	0
DB:	13	Gaps:	0
US-09-972-758A-2 (1-359) x BM129736 (1-557)			
OY	157	ProserlystLysLysArgHisTPrLysProTyrTyrLysLeuThrTPGluGluLysLys	176
Db	3	CCGTCCAAGAAAGAGCGGCGCATTTGGAACCCGACTACAACTGCAAGCTGGAGAGAGAGAAA	62
OY	177	LysPheAspGluLysGlnSerLeuArgAlaSerGllleArgAlaGluMePheAlaLys	196
Db	63	AAGTTTCACAGCAGAAACAGAGCTTCGAGCTTCAGAGATCCGAGCCGAGATGTTCCCAAG	122
OY	197	GlyGlnProValAlaIprOTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlu	216
Db	123	GGCCAGCGCGGCGGCGCTATTAACACGACGAGATTCTCATGATGATGATCAGACAGACAG	182
OY	217	GluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThr	236
Db	183	GAGCCGGAACCTCAAAACCGGCTCTACTCCAGCGGGCCCGCCGCAATCCGACAGACACC	242
OY	237	SerAspAspAspPheMetGluGluGlyGluGluLysAspGlySerAspGlyMetGly	256
Db	243	AGCGATGACGACTTCATGGAAGAGGGGCTGAGAGGATGGGGGCGACGATGGCATGGGA	302
OY	257	GlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGlnArgTyrHisThr	276
Db	303	GGGAGCGGACCGAGTTTCTGCACGGGACTTCTCGGAGACGTACGAGCCGTACACACG	362
OY	277	GluSerLeuGlnAsnMetSerLysGlnGluLeuLysGluTyrLeuGluLysLys	296
Db	363	GAGAGCTCGCAGAACATGAGCAACGACGAGCTCTCAAGAGATCTCGAACTGGAGAAC	422
OY	297	CysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGly	316
Db	423	TGCCTTCGCGCATGAGAGAGCAGAAACACCGCTCGCGCTCGAGACCAAGCGCTGGGT	482
OY	317	GlyAspAspAlaArgValArgGluLeuGlnGluLeuGlnAspArgLeuArgAlaGluAsn	336
Db	483	GGCCAGCAGACCCGCGTGTGCGGAGCTGAGCTGAGTGCAGCCGCGTGCGCCGAGAAC	542
OY	337	LeuGlnLeuLeuThr	341
Db	543	CTCCAGCTCGTCGACC	557
RESULT 15			
LOCUS	BF167480	762 bp	mRNA linear EST 30-OCT-2000
DEFINITION	60175078F01_NCL_CGAP_Lu29 Mus musculus cDNA clone IMAGE:395563 5', mRNA sequence.		
ACCESSION	BF167480		
VERSION	BF167480.1	GI:11047832	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 762)		
JOURNAL	NIH-MGC http://mgc.ncl.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: rga@bbs.fda.gov		
	Tissue Procurement: Gilbert Smith, Ph.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Invitrogen Genomics, Inc.		

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:23:40 ; Search time 48 Seconds
(without alignments)
799.640 Million cell updates/sec

Title: US-09-972-758a-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYQHQPOTSNCTG.....LTENELHROGERAPLSKRGD 359

Scoring table:
BIOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PCT05_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	359	9	US-09-972-758-2
2	1910	100.0	359	10	US-09-745-763-34
3	134.5	7.0	546	9	US-09-976-740-44
4	134.5	7.0	546	12	US-10-023-529-44
5	134.5	7.0	546	12	US-10-023-523-44
6	130	6.8	611	10	US-09-216-393-81
7	128	6.7	2476	9	US-09-824-574-7
8	124	6.5	489	10	US-09-876-889-350
9	123.5	6.5	557	9	US-09-976-740-5
10	123.5	6.5	557	10	US-09-962-055-5
11	123.5	6.5	557	12	US-10-023-529-5
12	123.5	6.5	557	12	US-10-023-523-5
13	121	6.3	538	10	US-09-827-822-9
14	120	6.3	427	9	US-10-050-704-273
15	120	6.3	483	9	US-10-050-704-272
16	120	6.3	911	9	US-10-245-103-52
17	120	6.3	911	9	US-10-245-107-52
18	120	6.3	911	9	US-10-245-143-52
19	120	6.3	911	9	US-10-245-771-52

20	120	6.3	911	9	US-10-245-851-52	Sequence 52, Appl
21	120	6.3	911	9	US-10-245-883-52	Sequence 52, Appl
22	120	6.3	911	9	US-10-237-553-52	Sequence 52, Appl
23	120	6.3	911	9	US-10-238-183-52	Sequence 52, Appl
24	120	6.3	911	9	US-10-238-283-52	Sequence 52, Appl
25	120	6.3	911	9	US-10-238-370-52	Sequence 52, Appl
26	120	6.3	911	9	US-10-245-055-52	Sequence 52, Appl
27	120	6.3	911	9	US-10-245-147-52	Sequence 52, Appl
28	120	6.3	911	9	US-10-245-730-52	Sequence 52, Appl
29	120	6.3	911	9	US-10-245-739-52	Sequence 52, Appl
30	120	6.3	911	9	US-10-246-210-52	Sequence 52, Appl
31	120	6.3	911	9	US-10-239-196-52	Sequence 52, Appl
32	120	6.3	911	9	US-10-243-024-52	Sequence 52, Appl
33	120	6.3	911	9	US-10-243-409-52	Sequence 52, Appl
34	120	6.3	911	9	US-10-245-033-52	Sequence 52, Appl
35	120	6.3	911	9	US-10-245-621-52	Sequence 52, Appl
36	120	6.3	911	9	US-10-245-880-52	Sequence 52, Appl
37	120	6.3	911	9	US-10-243-095-52	Sequence 52, Appl
38	120	6.3	911	9	US-10-245-185-52	Sequence 52, Appl
39	120	6.3	911	9	US-10-245-437-52	Sequence 52, Appl
40	120	6.3	911	9	US-10-245-473-52	Sequence 52, Appl
41	120	6.3	911	9	US-10-245-770-52	Sequence 52, Appl
42	120	6.3	911	9	US-10-245-877-52	Sequence 52, Appl
43	120	6.3	911	9	US-10-246-976-52	Sequence 52, Appl
44	120	6.3	911	9	US-10-243-320-52	Sequence 52, Appl
45	120	6.3	911	9	US-10-242-743-52	Sequence 52, Appl

ALIGNMENTS

RESULT 1
US-09-972-758-2
; Sequence 2, Application US/09972758
; Patent No. US20020160497A1
; GENERAL INFORMATION:
; APPLICANT: Case Western Reserve University
; APPLICANT: Montano, Monica
; APPLICANT: Wiltman, Bryan
; TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth
; FILE REFERENCE: 27708/04004
; CURRENT APPLICATION NUMBER: US/09/972,758
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/238,187
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 2
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-758-2

Query Match 100.0%; Score 1910; DB 9; Length 359;

Best Local Similarity 100.0%; Pred. No. 1.3e-122; Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAEPFLSEYQHQPOTSNCTGAANVOBELNPRPGAEERVPEDSRMOSRAFPOLGSRPG	60
DB	1	MAEPFLSEYQHQPOTSNCTGAANVOBELNPRPGAEERVPEDSRMOSRAFPOLGSRPG	60
QY	61	PEGGSLESPQPPLOTACPESSCLRGCEKQNGDDSSAGDDPPPAVEPTPAELLAQ	120
DB	61	PEGGSLESPQPPLOTACPESSCLRGCEKQNGDDSSAGDDPPPAVEPTPAELLAQ	120
QY	121	PCHDSEASKIGAPPAAGEEEMGQOQOLGKKRRRRRSKKRRHMKPYKLTWEKKKFE	180
DB	121	PCHDSEASKIGAPPAAGEEEMGQOQOLGKKRRRRRSKKRRHMKPYKLTWEKKKFE	180
QY	181	KOSLRASRIRAREMAKQPAVPYNTTQFLMDHDHQEERDLKTGLYSKRAAKSDTSDDD	240
DB	181	KOSLRASRIRAREMAKQPAVPYNTTQFLMDHDHQEERDLKTGLYSKRAAKSDTSDDD	240

QY 241 FMEEGEEDGSDGCGDSEFLORDFSTYERYHTESLQNNKOEELKEYELEKCLSR 300
Db 241 FMEEGEEDGSDGCGDSEFLORDFSTYERYHTESLQNNKOEELKEYELEKCLSR 300
QY 301 MEDENNRLRLLESKRLLGDDARVRELELELDRLRAENLQLTENELHROOERAPLSKFGD 359
Db 301 MEDENNRLRLLESKRLLGDDARVRELELELDRLRAENLQLTENELHROOERAPLSKFGD 359

RESULT 2
US-09-745-763-34

; Sequence 34, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

Lavallee, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 359 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Query Match 100.0%; Score 1910; DB 10; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.3e-122;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLSYOHQPTSNCTGAAYOEELNPERPRAEERVPEDDSRWQSAFPLQGRPG 60
Db 1 MAEPLSYOHQPTSNCTGAAYOEELNPERPRAEERVPEDDSRWQSAFPLQGRPG 60
QY 61 PEGGSLSESPPLQTOACPESSCLREGEKONGDSDAGDPPPAVEPTPEALLAQ 120
Db 61 PEGGSLSESPPLQTOACPESSCLREGEKONGDSDAGDPPPAVEPTPEALLAQ 120
QY 121 PCHDSEASKIGAPAGGEEWGOOOLGKKRRRRPSSKKRRHKPYKLTWEKKKFE 180
Db 121 PCHDSEASKIGAPAGGEEWGOOOLGKKRRRRPSSKKRRHKPYKLTWEKKKFE 180

QY 181 KOSLRASRIAMFAKGPVAPYNTQFLMDHDDEPDLKTLGYSKRAAAKSDPTSD 240
Db 181 KOSLRASRIAMFAKGPVAPYNTQFLMDHDDEPDLKTLGYSKRAAAKSDPTSD 240
QY 241 FMEEGEEDGSDGCGDSEFLORDFSTYERYHTESLQNNKOEELKEYELEKCLSR 300
Db 241 FMEEGEEDGSDGCGDSEFLORDFSTYERYHTESLQNNKOEELKEYELEKCLSR 300
QY 301 MEDENNRLRLLESKRLLGDDARVRELELELDRLRAENLQLTENELHROOERAPLSKFGD 359
Db 301 MEDENNRLRLLESKRLLGDDARVRELELELDRLRAENLQLTENELHROOERAPLSKFGD 359

RESULT 3
US-09-976-740-44

; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:

APPLICANT: Lees, Ann M.

Lees, Robert S.

APPLICANT: Law, Simon W.

APPLICANT: Arjona, Aubal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING

FILE REFERENCE: 10797-004001

CURRENT APPLICATION NUMBER: US/09/976,740

PRIOR FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: 09/616,289

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: US 08/979,608

PRIOR FILING DATE: 1997-11-26

PRIOR APPLICATION NUMBER: US 60/031,930

PRIOR FILING DATE: 1996-11-27

PRIOR APPLICATION NUMBER: US 60/048,547

PRIOR FILING DATE: 1997-06-03

NUMBER OF SEQ ID NOS: 53

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO: 44

LENGTH: 546

TYPE: PRT

ORGANISM: Homo sapiens

US-09-976-740-44

Query Match 7.0%; Score 134.5; DB 9; Length 546;
Best Local Similarity 21.9%; Pred. No. 0.15;
Matches 86; Conservative 51; Mismatches 134; Indels 121; Gaps 14;

QY 20 GAAVQELNPERPRAEERVPEDDSRWQSAFPLQGRPGPEGSLSESPPLQTOAC 79
Db 9 GAAVQELNPERPRAEERVPEDDSRWQSAFPLQGRPGPEGSLSESPPLQTOAC 79
QY 80 PEGGSLSESPPLQTOACPESSCLREGEKONGDSDAGDPPPAVEPTPEALLAQ 121
Db 80 PEGGSLSESPPLQTOACPESSCLREGEKONGDSDAGDPPPAVEPTPEALLAQ 121
QY 122 CHDSEASKL-----GAPAGGEEWGOOOLGKKRRRRPSSKKRRHKPYKLTWEKKKFE 180
Db 122 CHDSEASKL-----GAPAGGEEWGOOOLGKKRRRRPSSKKRRHKPYKLTWEKKKFE 180
QY 167 GLGKEITLIMQTLNLTSTPEEKLAALCKYALLLEHNSQOKMLLOKKQ-----SQ 219
Db 167 GLGKEITLIMQTLNLTSTPEEKLAALCKYALLLEHNSQOKMLLOKKQ-----SQ 219
QY 208 FLMDHDDEPDLKTLGYSKRAAAKSDPTSDDEMEGEGEDGSDGCGDSEFLORDF 267
Db 208 FLMDHDDEPDLKTLGYSKRAAAKSDPTSDDEMEGEGEDGSDGCGDSEFLORDF 267
QY 220 LV-----QEKDHLRGESKAVLARSK-----LESLREL 248
Db 220 LV-----QEKDHLRGESKAVLARSK-----LESLREL 248
QY 268 SETERYHTESLQNNKOEELKEYELEKCLSRMEDENNRLRLLESKRLLGDDARVRELE 327
Db 268 SETERYHTESLQNNKOEELKEYELEKCLSRMEDENNRLRLLESKRLLGDDARVRELE 327
QY 249 QR-----HNRLKKEGQVQRAEEERKEEYVSHQVLTNLTLOLOMEQHNRKNSKLRQENM 303
Db 249 QR-----HNRLKKEGQVQRAEEERKEEYVSHQVLTNLTLOLOMEQHNRKNSKLRQENM 303

CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-09-824-574-81

Query Match 6.8%; Score 130; DB 10; Length 611;
Best Local Similarity 22.2%; Pred. No. 0.34;
Matches 82; Conservative 58; Mismatches 155; Indels 74; Gaps 15;

QY 13 POTSCTGAAYQOELN-PPRPGAEERVP-----EDSRMOSRAF-OLGGRPGE 62
DB 189 PRFSKDYCCFQAAHLSLPPQSLGSSPSSPISVTMDYALFDSASPLHAGELSLPGAV 248
QY 63 GEGSLESQPP-PLQTAQPESSCLREGKQNGDDSSAGGDPPEPAVEPPEAEEL 118
DB 249 SASERLLTAPAEIGSASACLSVSC-----GPEMSPTADTTR- 287
QY 119 AOPCHDSASKLGAPAAAGE-----EWGQOOROLGKKHRRRPPSKKKHMKPYKLTWEE 174
DB 288 ---HDAEERERRRAEEERERERERERERERERERERERERERERERERERERER 339
QY 175 K---KKPFKOSLR-----SRIRAEPMFAKGPVAPYNTQFLMDHDOEEDPLTKGLYSK 227
DB 340 KAROEDEDERERRRVEEKKAROEDEDERERRRVEEKKARQ-----KEEERERRRVEEKK 395
QY 228 RAAAKSDTSDDEME-GEFEDGSDGNGGSDGSEFLORFSETYE--RYHTESLON 281
DB 396 AROEEERERRRVEEKKAROEDEDERERRRVEEKKAROEDEDERERRRVEEKKARQ 455
QY 282 MSKOELIKEYLELEKCLSMEDENRRLLESKRLGGDARVLELELDRLAENLQJLT 341
DB 456 REEERERRRVEEKKERERERERERERERERERERERERERERERERERERER 505
QY 342 ENELHROOE 350
DB 506 EKERERQEE 514

RESULT 7
US-09-824-574-7
; Sequence 7, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Mollanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Jone, Olli A.
; TITLE OF INVENTION: Arip4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 2476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-824-574-7

Query Match 6.7%; Score 128; DB 9; Length 2476;
Best Local Similarity 19.7%; Pred. No. 2.2;
Matches 78; Conservative 52; Mismatches 130; Indels 136; Gaps 13;

QY 43 EDSRMOSRAFPOLGPRPGE-GLSLESQPPLOTAQPESSCLREGKQNGDDSSA 99
DB 1142 EDNRKQKQRTSARAKKTKGTNTEKKRNLSRATPKRKQVITSSSSIDGDDONSAGESSD 1201

QY 100 GGFPP-----PAVEP-----TPE-----AELIA 119
DB 1202 EOKTKPYENLVPSHTGFCOSSGDEALSKSVPATVDDDDDDPENNIAKKMLLEELKA 1261
QY 120 QPCHDSASKLGAPAGEEEMGOO-----ROLGKKHRRRPPSKK- 161
DB 1262 NLSSDESDSDPEPGGKKRIRIGQSESPADDELRLREQLAVQNVNSESSDSESKP 1321
QY 162 --RHMKPYKLTWEEKKKFDEKQSLRASRIAEPMFAKGPVAPYNTQFLMDHDOEEDPD 219
DB 1322 RYRHLRLKHLTSLDGESEK-----PTKPEKKAKOR----- 1356
QY 220 LKTGLYSKRAAKSDTSDDFMEEGEGEDGSDGSGSEFLORFSETYERYHTESL 279
DB 1357 ---NRKVSSESDSDTDFQESGVSEVSE-----SEDDQRPRTSAKKALEEN 1402
QY 280 QMSKOELIKEYLELEKLSR-----MEENRRLLESKRL 315
DB 1403 QRSYKQKKRRRIKQVEDSSSEKNSHSEOKKEGDEDEDEDEDEDEDEDEDEDEDE 1458
QY 316 GDDARVLELELDRLAENLQJLTENELHROOER 351
DB 1459 GKGRKKIKTI-LKDDKLTTE-----TONALKEEER 1488

RESULT 8
US-09-876-889-350
; Sequence 350, Application US/09876889
; Patent No. US20020076715A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; FILE REFERENCE: 210121.466C3
; CURRENT APPLICATION NUMBER: US/09/876,889
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 353
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 350
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-876-889-350

Query Match 6.5%; Score 124; DB 10; Length 489;
Best Local Similarity 22.7%; Pred. No. 0.68;
Matches 71; Conservative 52; Mismatches 146; Indels 44; Gaps 12;

QY 7 SEYOHQPOTSCTGAAYQEE-----LNPER---PGAEERVP-EDSRMOSRAF 52
DB 3 SEYVSGDQ-SPDGCATPYQDERDSSGDGEDVYNEGHSSDPTGSYERHSNENTSDSENEEL 61
QY 53 POLGGRPGEGLSLESQPPLOTAQPESSCLREGKQNGDDSSAGGDPPEPAVEPT 112
DB 62 PK-----PRISDS-ESFDDP-RNQASDSEN--ELPKPRVSDSESGQKGPASDSPT 111
QY 113 PEAEELAOPCHDSASKLGAPAGEEEMGO-----QOROLGKKHRRRPPSKKKHMKPY 167
DB 112 DASRHQKQPESSDDSDRENK---GEDTEMONDSFHSDDHRRKFFHSDSSEEEHKKQ 167
QY 168 YKLTWEEKKKPFDEKQSLRASRIAEPMFAKGPVAPYNTQFLMDHDOEEDPLTKGLYSK 227
DB 168 MDSDEDEKEGEEKAYAKKAVALSDSEDEKASAKSKRVVSDADSDSDAVSKGKREK 227
QY 228 RAAAKSDTSDDFMEEGEGEDG--GSDGMSGDSEFLORF-----SETYERYHTE 277
DB 228 TIASDSEDEACKELSDKKNEKDLFGSDSESGNEENLADIFGSDGDEEEETFTGNQE 287
QY 278 SLQNSKOELIKE 290

```

Db      288 DLEEKGETQYKE 300

RESULT 9
US-09-976-740-5
: Sequence 5, Application US/09976740
: Publication No. US20020194633A1
: GENERAL INFORMATION:
:   APPLICANT: Lees, Ann M.
:   APPLICANT: Lees, Robert S.
:   APPLICANT: Law, Simon W.
:   APPLICANT: Arjona, Anibal A.
:   TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
:   TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
:   FILE REFERENCE: 10797-004001
:   CURRENT APPLICATION NUMBER: US/09/976,740
:   CURRENT FILING DATE: 2001-10-12
:   PRIOR APPLICATION NUMBER: 09/616,289
:   PRIOR FILING DATE: 2000-07-14
:   PRIOR APPLICATION NUMBER: US 08/979,608
:   PRIOR FILING DATE: 1997-11-26
:   PRIOR APPLICATION NUMBER: US 60/031,930
:   PRIOR FILING DATE: 1996-11-27
:   PRIOR APPLICATION NUMBER: US 60/048,547
:   PRIOR FILING DATE: 1997-06-03
:   NUMBER OF SEQ ID NOS: 53
:   SOFTWARE: FASTSEQ for Windows Version 4.0
:   SEQ ID NO 5
:   LENGTH: 557
:   TYPE: PRT
:   ORGANISM: Oryctolagus cuniculus
US-09-976-740-5

Query Match      6 54: Score 123.5; DB 9; Length 557;
Best Local Similarity 22.68; Pred. No. 0.85;
Matches 91: Conservative 53; Mismatches 123; Indels 135; Gaps 20;

QY 20 GAAVVOELNPERPGEAEERVP-----EDSRMQRAPFPLGGRPEGEGLSEQ 70
Db 9 GAA---KQPNKSSPGQPEAGAEAGQGRPGPARAREAGASSAPGR--PEGQAKTAQ 63
QY 71 PPPL-----OTQACPESSCREGEKGONGDDSSAGDFFPAVEPTPEALLAQP 121
Db 64 PGALCDVSESLRQLEDILSTYCV-DNNQCAPGEDVGQ-----EP-PEPE----- 107
QY 122 CHDSEASKL-----GAPAAEGDEEMGQO-----RQLGKKHRRPRSKKR 162
Db 108 --DAEKSRAYVARGEPERPGTPVYNCKETSKAEPTEELRTSDVGGRRRRPQEKKA 165
QY 163 HW-----KPYUKLTWEKKKFKDEKOSLRISRTIAEMFAKQGPAPAPNTT 206
Db 166 KGLCKEITLLMOTLINTLSTPEEKLAALCKYAEELLEHRNSQKQMKLLQKKQ-----S 218
QY 207 QFLMDHDOEPPDLTKGLYSKRAAKSDDN-----DDFMEEGEGEDGSDGMGD 258
Db 219 QLV-----QEKHLRGHSHKAILARSKLESCLRELQRHNSLKEDG----- 259
QY 259 GSEFLQRFSTTYRYRTYESLQNNKSKOELLKEYLEKCLSRMEDENNRLKESKRLGCD 318
Db 260 ---VQARAEEREEKRKVEYTSHFQTLNDI--QLQME---QHNEKRSKLROEMELA-- 306
QY 319 DARVRELELELDRLRAENL-----QLLTNELLHROE 350
Db 307 -ERLKKL-IEQYELREEHIDKVFHKKDLQOOLVAKLQAOE 346

```

```

1 Lees, Robert S.
2 Law, Simon W.
3 Arjona, Anibal A.
4 TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
5 BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
6 TREATING ATHEROSCLEROSIS
7
8 NUMBER OF SEQUENCES: 42
9
10 CORRESPONDENCE ADDRESS:
11 ADDRESS: Fish & Richardson P.C.
12 STREET: 225 Franklin Street
13 CITY: Boston
14 STATE: MA
15 COUNTRY: USA
16 ZIP: 02110-2804
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Diskette
20 COMPUTER: IBM Compatible
21 OPERATING SYSTEM: DOS
22 SOFTWARE: FASTSEQ for Windows Version 2.0
23
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/09/962,055
26 FILING DATE: 24-Sep-2001
27
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: 08/979,608
30 FILING DATE: 26-NOV-1997
31 APPLICATION NUMBER: US 60/031,930
32 FILING DATE: 27-NOV-1996
33
34 ATTORNEY/AGENT INFORMATION:
35 NAME: Myers, Louis
36 REGISTRATION NUMBER: 35,965
37 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
38
39 TELECOMMUNICATION INFORMATION:
40 TELEPHONE: 617/542-5070
41 TELEFAX: 617/542-8906
42
43 INFORMATION FOR SEQ ID NO: 5:
44
45 SEQUENCE CHARACTERISTICS:
46 LENGTH: 557 amino acids
47 TYPE: amino acid
48 TOPOLOGY: linear
49 MOLECULE TYPE: protein
50 FRAGMENT TYPE: internal
51 SEQUENCE DESCRIPTION: SEQ ID NO: 5:
52
53 US-09-962-055-5
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949

```



```

: CURRENT FILING DATE: 2001-04-06
: PRIOR APPLICATION NUMBER: US/09/370,368
: PRIOR FILING DATE: 1999-08-09
: NUMBER OF SEQ. ID NOS: 10
: SOFTWARE: FASTSEQ for Windows 3.0
: SEQ. ID NO. 9
: LENGTH: 538
: TYPE: PRT
: ORGANISM: Moloney Murine Leukemia Virus
: US-09-827-822-9

```

Query Match	6.3%	Score 121	DB 10	Length 538
Best Local Similarity	21.3%	Pred. No. 1.2		
Matches 86	Conservative 49	Mismatches 147	Indels 122	Gaps 19

[illegible]

```

RESULT 14
US-10-050-704-273
? Sequence 273, Application US/10050704
? Publication No. US20030050442A1
? GENERAL INFORMATION:
? APPLICANT: Ruben et al.
? TITLE OF INVENTION: 62 Human Secreted Proteins
? FILE REFERENCE: P2039P1
? CURRENT APPLICATION NUMBER: US/10/050,704
? CURRENT FILING DATE: 2002-01-18
? PRIOR APPLICATION NUMBER: 09/684,524
? PRIOR FILING DATE: 2000-10-10
? PRIOR APPLICATION NUMBER: PCT/US00/08979
? PRIOR FILING DATE: 2000-04-06
? PRIOR APPLICATION NUMBER: 60/128,693
? PRIOR FILING DATE: 1999-04-09
? PRIOR APPLICATION NUMBER: 60/130,991
? PRIOR FILING DATE: 1999-04-26
? NUMBER OF SEQ ID NOS: 344
? SOFTWARE: Patentln Ver. 2.0
? SEQ ID NO 273
? LENGTH: 427
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-050-704-273

```

Query Match	6.38	Score	120	DB	9	Length	427
Best Local Similarity	24.48	Pred. No.	1.1				
Matches	98	Conservative	35	Mismatches	141	Indels	128
				Gaps	22		

```

OY      6 LSEYOH--PQT--SNCTGAARVOEL--NPEPPAERVRVEDSRMOSRAFPOLCG 57
Db      72 LLEFEFQSMVPPGFSEEGKALEEEKEEDEEKEEEDEEKEEVEDEALW---AMPSELS 128
OY      58 RPGEGRGSLSEOP-----PPLQOQACSESCLEKEGCKGONGDDSSAGDPPPAE 108
Db      129 SPQC---EASLPTPAQOEESLSQARAKAVLOPGASPLPDGESEASRPVRHG---PPTTE 182
OY      109 VEPTPEAELLQPCHS--EASKLGAAPAGE-----EEAGQOOROLGKKKRRRP 157
Db      183 TLPTPRRNLASPSPTLEAREVEGATNGPCLSGVRGSESEELTGSSGAPSLLPATRAP 242
OY      158 SKKRHKMPYKLTWEKKKFKPEKOSLRASRTIABEMFAKGPVAPVNTTQFLMDHDQEE 217
Db      243 EGTRELEAP-----SEDNSGR-----TAPAGTSV-----QAO 269
OY      218 PDLKTGLYSKRAAKSDTSD--DDFMEEGG--EEDGG-----SDGAGGP-----GSEFL 263
Db      270 PVLPTDASAKGVAVVPASGDCVPSPCHNGCTCLEEEGVARCILCPGIGGGLCDVGRFPC 329
OY      264 Q-----RDFS-----EYERY--HTESLONNSKQELIKVLELEKCLSRM 301
Db      330 NPGMDAFQOGACYHFFSTRSMEEAEQCMMCAHLASTIEPEODFI----- 376
OY      302 EDENNRULESKRGLGDDARVRELELELDRLAENMLQLTEN 343
Db      377 ---NNRYR-EYOWIGLNDRTI---EGGFLMSDSDVPLLYEN 409

```

```

RESULT 5
US-10-050-704-272
: Sequence 272, Application US/10050704
: Publication No. US20030050442A1
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: 62 Human Secreted Prc
: FILE REFERENCE: P2039Pl
: CURRENT APPLICATION NUMBER: US/10/050,704
: PENDING FILING DATE: 2002-01-18
: PRIOR APPLICATION NUMBER: 09/684,524
: PRIOR FILING DATE: 2000-10-10
: PRIOR APPLICATION NUMBER: PCT/US00/08979
: PRIOR FILING DATE: 2000-04-06
: PRIOR APPLICATION NUMBER: 60/128,693
: PRIOR FILING DATE: 1999-04-09
: PRIOR APPLICATION NUMBER: 60/130,991
: PRIOR FILING DATE: 1999-04-26
: NUMBER OF SEQ ID NOS: 344
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 272
: LENGTH: 483
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-050-704-272

```

Query Match	6.38	Score 120	DB 9	Length 483
Best Local Similarity	24.48	Pred. No. 1.2		
Matches 98	Conservative 35	Mismatches 141	Indels 128	Gaps 22

Oy 6 LSEYQHO----POT--SNCTGAAYQDEL--NPEPPAAERRYVEDDSRWQSRAPFOLGCG 57
 Db 124 LLEFFQSWMPPTGFGSEEGKALFEDEEKRYEDEEKEEEDDEEVEDEALW---AWPSELTS 180
 Oy 58 RQPEDEGSLLESP-----PPLQYQACSESCLEBKGKQNGCDDSSAGGDFPPPAE 108
 Db 181 SPSP--EASLPTEPPAAQSESLSQAPARAVLQPGASPLPDGSEASRPVHG----PPTTE 234
 Oy 109 VEPTEALLAOPCHDS--EASKLPAAPAGF-----EAGGQOQROLGKKKHRRP 157
 Db 235 TLPPPRRNINASTPSTLVEAREVECATGPPSLGVPGESEETFGSGEAGSLLPATRAP 294
 Oy 158 SKKRHKMPYKLTWEKKKPFQKSLRASRIIAPEMFQKQVPAPYNTTQFIMDDHOEF 217

```
Db      295 EGTRLEAP-----SEDNSGR-----TAPAGTSV-----OAQ 321
QY      218 PDLKTGLYSKRAAKSDPTSD--DDFMEEGG--EEDGG-----SDGMGGD-----GSEFL 263
Db      322 PVLPTDSASRGVAVVPASGDCVPSPCHNGTCTLEEBEGVRCCLPGYGGDLCDVGLNFC 381
QY      264 Q-----RDFS-----ETERY--HTESLQNMKSQELIKEYLELEKCLSRM 301
Db      382 NPGMDAFOGACYKHFSRSTRSWEAFTOCRMVGAHLASISTPEODFI----- 428
QY      302 EDENNRRLRESKRIGDDARVRELELDRLRAENLQLTEN 343
Db      429 ---NNRYR-EYOWIGLNDRTI---ESDFLMSDGVPLLYEN 461
```

Search completed: June 17, 2003, 10:33:41
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:17:20 ; Search time 41 Seconds
(without alignments)
841.762 Million cell updates/sec

Title: US-09-972-758A-2

Perfect score: 1910

Sequence: 1 MAEPFLSEYHQHPQTSNCTG.....LTENELHROGERAPLSKFGD 359

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	148.5	7.8	992	2	T46337
2	139	7.3	755	2	S32103
3	134	7.0	568	1	I61106
4	133	7.0	462	1	A60746
5	133	7.0	538	1	FOVM1M
6	133	7.0	1493	2	A44224
7	132.5	6.9	530	2	JC7168
8	131	6.9	681	2	E5436
9	131	6.9	1095	2	T00329
10	131	6.9	1898	1	A45973
11	131	6.9	3488	2	T34418
12	129	6.8	1054	2	T30177
13	128	6.7	1733	1	B45344
14	127.5	6.7	852	2	T06310
15	127.5	6.7	918	2	A88188
16	127.5	6.7	4687	1	A39638
17	126.5	6.6	777	2	F54024
18	126.5	6.6	1407	2	S28589
19	125.5	6.6	779	2	E54024
20	125.5	6.6	1549	1	A40691
21	125.5	6.6	1909	2	A45592
22	125	6.5	383	2	C96581
23	125	6.5	913	2	T52485
24	124.5	6.5	474	2	T16441
25	123.5	6.5	677	1	S09078
26	123.5	6.5	1870	2	S37671
27	123.5	6.5	1872	2	S36152
28	123	6.4	723	2	A48217
29	123	6.4	1829	2	T41751

30	122	6.4	705	2	S32644	nucleolin - Africa
31	122	6.4	723	2	S33688	hypothetical prote
32	121.5	6.4	1974	2	T30010	hypothetical prote
33	121	6.3	538	1	FOVM1M	gag polyprotein -
34	120.5	6.3	849	1	S64732	scarfold attachmen
35	120	6.3	536	1	FOVM1M	gag polyprotein -
36	120	6.3	138073	6	S18073	nucleolar phosphop
37	120	6.3	1325	2	S16129	dynein-associated
38	120	6.3	1663	2	T42092	s-afadin - rat
39	119.5	6.3	451	2	C70241	hypothetical prote
40	119.5	6.3	651	2	F86563	hypothetical prote
41	119.5	6.3	651	2	A72060	hypothetical prote
42	119.5	6.3	1233	2	T14157	serine/threonine p
43	119.5	6.3	1560	2	T30282	calcium-binding pr
44	119	6.2	449	1	A41520	chromogranin A pre
45	119	6.2	655	2	T28885	hypothetical prote

ALIGNMENTS

RESULT 1

T46337 hypothetical protein DKFZp43402413.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46337

R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, January 2000

A:Reference number: 223037

A:Accession: T46337

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-992 <AAA>

A:Cross-references: EMBL:AL137265

A:Experimental source: adult testis; clone DKFZp43402413

C:Genetics:

A:Note: DKFZp43402413.1

Query Match	Score	148.5	DB 2	Length	992
Best Local Similarity	21.1%	Pred. No.	0.19		
Matches	94	Conservative	66	Mismatches	149
				Indels	137
				Gaps	17
OY	25	OELNPPPPGAEVVP	-----EEDSRWQSRAF	-----POLGPRPCEG	-----64
DB	230	ODELQSKQSKGLEELSPPLPHEERAQSPPSLATFEPPQ	-----GPDGPEKKEAEL	-----	284
OY	65	-----GSLEQPPPLQTA-CPESSCLREGKGGDDSSAGDFPPPAE	-----VEPTP	-----	113
DB	285	GEDSAASLSLQSLQREDQAPSPACCEKKGQHSQAEELGPGQEAEDPEEKVAVSPTRP	-----	-----	344
OY	114	-----EAEFLAQPCHDSEASKLGAAGGEEBKQDQROLGKK	-----	-----	152
DB	345	VSPEVRSTPEVAPPPOLSEAA-LKAMEAVAVQVLEDDQRLHLESKQEMQOLREKLCOE	-----	-----	403
OY	133	-----HRRRPSKKRHKMPYKLTWEKKKPFDEKQSLRASRTAEKFAFGQPPAVNT	-----	-----	205
DB	404	EELILRLHQKQSLSLRERLQKAI EEEEMRREESQRLSWLAQVOOSTQA	-----	-----	457
OY	206	TQFLMDHDEEPDLTKGLYSKRAAKSD	-----DTSDDPFMEEGEGDGS	-----	252
DB	458	-----DDQIRAEQASLQKREELQSQAKAERASLQKKNQMLQKLEELKSEKSEQ	-----	-----	511
OY	253	-----DGMGGDSFLLQDFSETYRY	-----HTE	-----SLQNSK	284
DB	512	AALNAKKEKALQQLREQLGEEKKEVAVTLERHLSAEELRLCSLEAKHREAVSSLOKRIQ	-----	-----	571
OY	285	OELIKVELEKCSLSEMEDENNRLRLSKRLGGDARVREL	-----ELEDRL	-----	332
DB	572	EAQKKEAQLOKCLQGV	-----HRVQKSTHYAGTVEHLSLIREKROVEGBEHERLDKM	-----	628
OY	333	RAENLQLLT	-----ENELHROGERAPL	-----	354
		:	:	:	:

Db 629 KEEHQVMKARQYAEERKORAEI 654

RESULT 2

S32103

filensin - bovine
N:Alternate names: intermediate filament protein

C:Species: Bos primigenius taurus (cattle)

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #ext_change 10-Sep-1997

C:Accession: S32103; A40690

R:Gounari, F.; Merdes, A.; Quinlan, R.; Hess, J.; Fitzgerald, P.G.; Ouzounis, C.; Georga

submitted to the EMBL Data Library, March 1993

A:Description: Bovine filensin possesses primary and secondary structure similarity to

A:Reference number: S32103

A:Accession: S32103

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-755 <GOU>

A:Cross-references: EMBL:X72388; NID:g287751; PID:g287752

R:Gounari, F.; Merdes, A.; Quinlan, R.; Hess, J.; Fitzgerald, P.G.; Ouzounis, C.A.; Geor

J. Cell Biol. 121, 847-853, 1993

A:Title: Bovine filensin possesses primary and secondary structure similarity to interme

A:Reference number: A40690; MUID:93260017; PMID:8491777

A:Accession: A40690

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-622, RP, 625-755 <GOU>

A:Experimental source: lens

A:Note: sequence extracted from NCBI backbone (NCBIN:132495, NCBIP:132499)

C:Keywords: membrane-associated protein

Query Match 7.3%; Score 139; DB 2; Length 755;

Best Local Similarity 22.6%; Pred. No. 0.52;

Matches 88; Conservative 32; Mismatches 117; Indels 152; Gaps 16;

26 EELNPRPGAEVRVEEDSRMOSRAFPOLGHRPREGSLESOP-PLQTOA----- 78

414 EASPTQOGAEDVP--DGRKISKAFELKMKIKVKPKPEPPADLYTKGRVYVS 471

79 -----CPSSCLREGKEKGQNGDSSAGDPPPAVEPTP-----EATLLAO 120

472 GGGSEYDFPCYFVSYPKAGGVVSKGDDS-----VPPDSGVPSPOQPPPLEEGGPPQ 526

121 PCHDSEASKIGAPAAAG-----EEWG 142

527 EKEDGLKEGCGPEPGKGDGVKKEGGPEPGKGDGVKKEGGPEPGKGDGVKKEGG 586

143 QOO-RQLGKKHRRRRSKKRRHKKPYKLTWEKKKFFDEKOSLRASRIRAEFAKQPYA 201

587 PEGKGDGVKKEGPEPGKEGLK-----EEGFLQEKED-----GQSPY 626

202 PYNTQFLMDHDDQEPDLKTGLYSKRAAKSDTSDDFMEEG-----GEED-- 249

627 PHPA-----DKQDEKAKELKGLQKQ-----DQKEEGAGCPMPWAPGEGS 671

250 -----GSDMGSGDSEFLQ-----DFSETYRYHTESLQNSKQPL 287

672 TPDSQPOVILGSEHGARGSGSRLASPPRIAYEKVYMESIEKFSIESIQTYETAV 731

288 IKEYLELEKCLSRMEDENNRLRESKRLG 316

732 IVETM-IEK-----TRANKKILG 748

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

A:Title: The involucrin genes of the mouse and the rat: study of their shared repeats

A:Reference number: A49377; MUID:94104476; PMID:8277848

A:Accession: I61106

A>Status: translation not shown; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-568 <RES>

A:Cross-references: GB:I28818; NID:9454427; PIDN:AAA1445.1; PID:9454428

C:Comment: During the terminal differentiation of keratinocytes, this protein from th

linked envelope under the plasma membrane.

C:Superfamily: involucrin

C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat

Query Match 7.0%; Score 134; DB 1; Length 568;

Best Local Similarity 20.9%; Pred. No. 0.76;

Matches 93; Conservative 66; Mismatches 169; Indels 118; Gaps 18;

3 EPLSEYCHONQOTSNTGAAVQEE-----LNPERPGAEKVRPEED 44

127 EQLRHQHQPOQESQGLGLQOQDVLAPQELHMQHOKELQDEPLPLGQOQKTPEDQ 186

45 -----SRWQRAFPOLGRRPE-----GEGSLESOPPLQTOAC-----PESSCLR 86

187 ELLGKQKQLHLYENRQPORQDELHNGQKQOQPOQEBLDLYQHQKQHNPE-LCLR 245

87 EEBKQNGDSSAGDPPPAVEP-----TRAEALLAQPCHDSEASKLGAP 133

246 KQOQOESHERELHLGKQOQOESHERELHLGKQOQOESHERELHLGKQOQOESCEPEL 302

134 AAGGEEMQOQORO-----LGKKHNR-----RPSKKRRHKKPYKLTWEKKRP- 178

303 HLGEOQHOBQOQOESCEPELHLGKQOQOESCEPELQKQKPHEDWLPDQKQKHL 362

179 -----DEKQSLRASRIRAE-----FAKQAPAPYNTQFLMDHDDQEP 218

363 DEPLHNGQOQOESQESLQVEKQKQESPEPELHNGQOQELHEDPMDQEKQSLHNP 422

219 DKTGLYSKRAAKSDTSDPPMERGEGEDGSGDGGSEFLQDFSEYERYHTES 278

423 ELHLG-----KQOESHEPMDTEQKQSLYEPELHLGKQOQOELYEQYR--SKS 472

279 LQNSKQOELIKELYLEKLSRMEDENNRL--RLSKRLGQDANVRELELDLRLAEN 336

473 LQQLKQEKASRGQELDD--SHLEQKELLQDLQDELVNKDE-----QLERKKHKL 524

337 L-----QLLTENELHROQERAPL 354

525 LQKQEKQIKQLVPSPD--RVQETQPI 548

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

R:Hutton, J C', Nielsen, E.: Kasten, W.
PERS Lett. 236, 269-274, 1988
A>Title: The molecular cloning of the chromogranin A-like precursor of beta-granin and P
A:Reference number: S02543; MOID:88312980; PMID:3044825
A:Accession: S02543
A>Status: nucleic acid sequence not shown: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 19-107,'0000',108-462 <HUT>
A:Note: part of this sequence, including the amino end of beta-granin, was determined by
C:Superfamily: chromogranin A
C:Keywords: glycoprotein; phosphoprotein
F:1-18/Domain: signal sequence #status predicted <SG>
F:19-462/Product: chromogranin A #status predicted <Mat>
F:92-107/Product: glutamine-rich
F:417-419/Region: cell attachment (R-G-D) motif
F:35-56/Disulfide bonds: #status predicted
F:68,147,216,349,358,366,408,409,413/Binding site: phosphate (Ser) (covalent) #status pr
F:189/Binding site: carbohydrate (Asn) (covalent) #status predicted

	Query Match	7.08;	Score 133;	DB 1;	Length 462;
	Best Local Similarity	24.2%;	Pred. No. 0.71;		
	Matches	87;	Conservative	54;	Mismatches 159; Indels 60; Gaps 17;
Qy	3	EPFLSYOHOPQTSNCTGAAYVOEELNPERPRGAEERPEEDSRMSRAFPQLGGRRPGE	62		
		: : :	:	: : :	:
Dd	64	EYVLSILRIHQ-----NLKLEODDLALOGAKERRAQQQQQQQQQQ--QDQQHSSFE	112		
Qy	63	GEGS--LESQPPPLQ---TQACPRESSCLREGEGAGONGDDSDAG--FPP-PPAEVEPRP	113		
		: : :	:	: : :	:
Dd	113	DELSEVFENQSPRAKHGDAASEAAPSKDTVERREDSKDQQQAFTETGCPRQAFPREPQ	172		
Qy	114	EAEELL--APPCHESEASKLGPAPAAGEEENGQOOROLKKHHRRPS--KKKRHKPKYYK	169		
		: : :	:	: : :	:
Dd	173	ESSMMGNSSQPGEDTANNTQSPTSLPSEQEHGIPQTTEGSE--RQPSAQQQARAKKOEK	229		
Qy	170	LTFWEKKRKFKDEKSLRASIRIAEMFAKGOPYAPAYNT-----TQFLMDHDDEPDCLKG	223		
		: : :	:	: : :	:
Dd	230	EEEEEEKKEEEEREKKAIRAKAGKPEVPFAASSHPYSCTYKIKQKDDGQSESOAYNG	289		
Qy	224	LYSKRAAASDDTDSDDFMEEGEEDGSDGM-----GCDSFEFLQRDFSITYERY	274		
		: : :	:	: : :	:
Dd	290	--KTGSAAVAPSECKGELEHSQOEEDEGEAMAGPQCGLPFGKGOE--LEHQQEEEEE--	343		
Qy	275	HTEESLONMKSQEILEFYLELEKLRSMEDENNRLFSKSLRGCGDDARYRELLELDRLRA	334		
		: : :	:	: : :	:
Dd	344	-----DEERSLRPW--EDKRWMSRMDOJLAKELTAAE-KRLEGEDDPDRSMKLSF-RARA	391		

RESULT 5
FORM1M
gag polyprotein - Moloney murine leukemia virus
N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein p16
C:Species: Moloney murine leukemia virus
A:Note: host Mus sp. (mouse)
C:Date: 01-Sep-1981 #sequence_revision 27-Nov-1985 #text_change 24-Jul-1997
A:Accession: A03930
R:Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.
Nature 293, 543-548, 1981
A>Title: Nucleotide sequence of Moloney murine leukaemia virus.
A:Reference number: A93265; MUID:82035843; PMID:6169994
A:Accession: A03930
A:Molecule type: genomic RNA
A:Residues: 1-538 <SHI>
A:Experimental source: clone PMLV-1
C:Comment: This protein is synthesized as a gag-pol polypolyprotein.
C:Genetics:
A:Gene: gag
C:Superfamily: mammalian retrovirus gag polyprotein I
C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein
F:2-131/Product: core protein p15 #status predicted <CI5>
F:132-215/Product: inner coat protein p12 #status predicted <CI2>
F:216-478/Product: core shell protein p30 #status predicted <C30>
F:479-534/Product: nucleoprotein p10 #status predicted <CI0>

Query Match	7.0%;	Score 133;	DB 1;	Length 538;
Best Local Similarity	20.9%;	Pred. No. 0.83;		
Matches	89;	Conservative	40;	Mismatches 142;
				Indels 154;
				Gaps 19;

```

QY 30 PERPPGA-----EERVPEEDSRKWSRAFPOLGGRPE---GEES-----LESOPPL 74
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 107 PPLPSAPSLPLEPPRSTPPRSSLYPALP-PSLGAKPKPOVUSDGGLIDLTEDDPV 165
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 QTAQPESSCLERGEKGCGDDSSAGGDPPEAPVEVPTPEAEELL-----AOPCHDSASKL 130
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 RPPRRPPSD -KDG-----NGCEATPAGEAR-----DPSEMASLRRRRPREVADSTTSSA 214
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 131 GAPAAGEEEMWGOOROLGKKKKRRRRSSKKRRHKRPYLYLWEKKKKPEKOSLRASRLR 150
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 PPLRAGGN-----GOLQWPFSSSDLYNWK-----NNNSFSEDPGKLTALIE 257
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 191 AEMFAKGGVAVNYNTQFLMDHDDEEPLDKTG-----LYSKRAAKSDD-----235
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 258 SVLTI-----THQPIWDDCCOQLGTLTIGEKKOYVLEARKKAVRGDGRFPOLPNE 307
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 236 -----TSDDDPMEEGEE-----DGSDDMGGDSE 261
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 308 VDAAPFLERPDMDDYTQAGRNHLVHYROLLAGLOAGRSPTNLAKVKGITQCPNESP5A 367
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 FLQRPFSLEYEK-----HTESLQNSMKQELIKE 290
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 368 FLER-LKAVRYRTPYDPEDPGQETNVSM5FTWQ5APDIGRKLGRLEDLKNKLGDLVR- 425
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 291 YLEKCLSRMDENNRRLRLSKSLRGGDARV-RELELELDRLRAVENLOLTENELHQO 349
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 --EAKIKNKRTPEER-----EERIRRETEKEERKRTTEDQEKERDRRRIR 472
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 350 ERAPL 354
      | | |
Db 473 EMSKL 477

```

RESULT 6
AA4224
DNA repair helicase ERCC6 - human
C:Species: Homo sapiens (man)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: AA4224
R:Torresblanca, C.; van Gool, A.; de Wit, J.; Vermeulen, W.; Bootsma, D.; Hoeijmakers, J.
Cell 71, 938-953, 1992
A:Title: ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne
A:Reference number: AA4224; MUID:93092214; PMID:1339317
A:Accession: AA4224
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1493 <TRO>
A:Cross-references: GB:U04791; NID:q182180; PIDN:AAA52397.1; PID:q182181
A:Note: sequence extracted from NCBI backbone (NCBIP:120238)
C:Keywords: DNA binding

	Query Match	Score	DB 2:	Length
Best Local Similarity	7.0%;	133;		1493:
Matches	Conservative 23.7%;	Pred. No. 2.3;	Mismatches 118;	Indels 104; Gaps 17
Oy	35 GAEEFVP-----EEDSRMOSRAFPQLGRPPEGEGLSESPPLOIQAQCESSCLREG	88	::	
Dd	201 GAEXKIELDHASLEDA-----EPDP---SLSGLMLMVQETAME--LIRTG	243	::	
Oy	89 EKGNGDDSSAGGDFPPAEVEPTP---EA-----ELLAOPCHDSEAK	129	:: ::	
Dd	244 QM-----TPROTOIPQOKKKPRKIMLNENASGFELYLAODAKLSFERKKOCCNRARAK	297	: ::	
Oy	130 LGAPA-----AGSEEMWGOQOROLKKNNRRPKSKKNHMRPYUULLTDEKKKPFKEKSL	184	::	
Dd	298 APAPVTPRAYONKKRPPKKAIVLSKKEER----LKHHIKKLQKRALQFOGVGLPAR	352	::	
Oy	185 RA--SRIAEPMFKSGCPVARNTYTLAMDHDNE---EPDL-----KTG	223	::	

```

Db      353  RPWESDMRPEABEDSGEE-SEYFPTBEEBEEDEDEVBGADJLDSGDTYELKPLPKGG 4111
QY      224  LYKRAAASDDTSDDDFMEEBEGEE-DGSGDGMGDSSELIÖRPFSEYRYHTESJÖNN 2822
Db      412  KRÖKKVPVÖE---IDDDFFPSSGEEAEAAVSEGGGGGRKKRGRRDDGDEDYVYKORLRW 4688
QY      283  SKÖELIKEYLELEKCLSRMEDENNRLRESKRLGSDDARVARELE 328
Db      469  NK-----LRLÖDKERKRULE-----DDEESDAEDF 494

```

```

RESULT 7
JC7168
lens epithelium-derived growth factor - human
N:Alternate names: transcriptional co-activator p75 protein
C:Species: Homo sapiens (man)
C:Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 11-May-2000
#accession JC7168

```

Accession: J07168

A:Molecule type: mRNA
A:Residues: 1-530 <SIN>
A:Cross-references: GB:AF063020
A:Experimental source: lens epithelial cells
A:Comment: This protein, a regulatory factor, belongs to a family of growth factor, which is involved in the regulation of this factor stimulates cell attachment, shortens the initial lag in growth, and e
C:Genetics:
A:Gene: ledgf
A:Keywords: eye lens; glycoprotein; growth factor; growth regulation

Query Match	6.9%	Score 132.5;	DB 2;	Length 530;
Best Local Similarity	21.2%	Pred. No. 0.87;		
Matches 77;	Conservative 56;	Mismatches 115;	Indels 115;	Gaps 17;

```
Oy      7 SEYOHQPOTSNCCTGAAYOEELINPERPPGAEERYEEDSRMQSRAFPQ-----L 55
```

55 GCPBQ--DECFESLFSQBBDIOTAOABDFSSCI BCFCEKCONNNDCAGCCDDBDPAUV--E 110
56 330QAAATNQSMASSDVEVEEA-----EISVSAEDIEDHEKASNEDYINAVDITTPK 143

```

Db      144  AARGRRKAKEQVETEAGVTTATASVNLKVPKGR-----PATEAKIPK 192
          | | : | : | : | : | : | : | : | : | : | : | : |

```

Qy 111 PTPAEELIAQPCDSEASKLGAIPAAGGEEMGQQQRLGKKKHRRRPSKKKRWKPYKL 170
::: ||| ::::

171	ТРЕБОВАНИЯ К БАЗИСНЫМ МОДУЛЯМ	229
193	ПРОГРАММА РС-РЕСДИ	229
200	ТРЕБОВАНИЯ К БАЗИСНЫМ МОДУЛЯМ	229

```

Db      230  ---:  ||  :  |  |
          QPKDDEGQK-----:  ||  :  ||  |  |  |
          -----EEDKPRKEPDKKEGKKEVESK 260

```

Qy 228 RA-AAKSDTSDDDFMEEGEEDGSDGMGSDSEFLORDSEFYER-----YHTESLQ 280

DB 261 RKNLAKTGVTSISDSSEESDDQJESKKRKGG -----RNFQYAHRRNRNLKGQHEKAEAD 313

```

314 RRRQF---EOMETQ---ONKDECKPPEVKKVEKKEHSDMSDLPORIHAEIKNSIKIDN 367
      ||| | : | : : : | : : | | : | : | : | : | : | : | : | : |

```

QY 337 LQL 339

Db	368	LDV	370
----	-----	-----	-----

RESULT 8
E59436

KIAA1314 protein [Imported] - huamn
C:Species: Homo sapiens (man)

C:\Accession: E59436; F59436

R:Nagase, T.; Kikuno, R.; Ishikawa, K.I.; Hirosewa, M.; Ohara, O.
DNA Res. 7, 65-73, 2000
A:Title: Prediction of the coding sequences of unidentified human genes. XVI. The com
A:Reference number: E59436

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-681 <NNG>
A:Cross-references: GB:BAAG2552; PID:G7243009; PIDN:BAAG2552.1
A:Ohara, O.; Nagase, T.; Kikuno, R.
Submitted to Genbank, January 2000
Reference number: F59436

A:Accession: F59436
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-681 <NMG>
A:Cross-references: GB:BAA92552, PID:g72a3009; PTDN:BAA92552.1.

Query Match	6.9%	Score 131	DB 2	Length 681
Best Local Similarity	26.4%	Pred	No. 1.4	
Matches	84	Conservative	44	Mismatches 110; Indels 80; Gaps 18

```
QY      66 SLESQPPLQIQADESSC-----LRGKGGONGDSSAGG-DfPPPAVEPTPEAE 116
        | | | | : : | | | | | | | | | | | | | | | | | | | | | |
Db      14 SNESIHHPAERSNSAESVDSASMEDFWKEIEISIK---DSMGCGGEPPPAVEPTPYDEGE 70
```

[illegible]

177 KFDEKOSLRASIRAEWFAKQ----PVADYNTQFELMDHDHQEPPD----KTGLYSK 227

```

Db      122 K-DKQSI RDR---DIEGVSESPPRDTCGHTNQL---DGIMKEERELPRVIKTS GMPD 173

```

```
QY      228 RAAASDDTSDDFMEEGSEEDGSGDGM-----GGGS-----EF 262
```

263 LÖRD-----FSEYERYHTESLÖ--NNSKOELIKEYLELEKTSRMEDENNRLRL- 310

Db 234 LEKNIPPEELSFVSYSEMVTBALKRNLKSKSEIKKE-----DYVLTGFNVQKTRFGILT 289

QY **311 ESKRIGDD-ARVRELEL 327**

biochemical) protein kinase - kinase (assay)
T00329
RESULT 9

C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000

R. Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; G. DNA Res. 5, 31-39, 1998

Title: Prediction of the coding sequences of unidentified human genes. IX. The
 A: Reference number: 214086; MUID:98290545; PMID:9628581

A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA

A: residues: 1-1095 <NAG>
A: Cross-references: EMBL:AB011125; NID:g3043629; PID:g3043630
A: Exported from source: bris1

C:Genetics:
A:Note: KIAA0553

Query Match	6.9%	Score 131;	DB 2;	Length 1095;
Best Local Similarity	19.7%	Pred No. 2	2.	

Matches 99; Conservative 41; Mismatches 137; Indels 226; Gaps 21.

Db 33 PKPKSCIKAAASO-----CAEKTVEVSEOPKETSMPER--SEPGSKAEAKKALG 80

submitted to the EMBL Data Library, July 1996
A:Description: SL12 homologue of Yarowia lipolytica.
A:Reference number: 220763
A:Accession: F30177
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1054 <GAD>
A:Cross-references: EMBL:U65409; NID:93978133; PID:93978134; PIDN:AA083184.1
A:Experimental source: strain W29
C:Genetics:
A:Gene: SLA2
A:Map position: III, adjacent to MATA

Query Match 6.8%; Score 129; DB 2; Length 1054;
Best Local Similarity 23.4%; Pred. No. 2.8;
Matches 92; Conservative 45; Mismatches 124; Indels 132; Gaps 19;

```
31 ERPPGAEERVPEDSRMSRAFPOLGGRPEGE-GSLESOPPLQTOACESSCLREG 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 KNPNLLE---EEDG-----PGLPQPKSVAEIRSVSPAPLPV-ATPTPSI---- 305

QY 90 KGQNDSSAGDPPPAEVEPT---PEALLAQPCHDSASKLGAFAAGGEEMQ-- 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 -----PAEAPINFMQDALLAQOQYDAQERLRQQAHEERIRMQ 349

QY 144 -QOROLGKKHRRPSSKKRHKPYKLTWEKKKFE---KSLRASRIAEAFKQ 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 350 MQOQOQFEMOQROQMEAOQRA-----OEOIMADQMARHAGRMALERDLALRG 400

QY 199 PVAAYNTQFLMDHDOEPRLTGKLYSKRAAASDOTSDDFMEEGEGEGSDMGD 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 ---YDDQQLMEYDRKVALLEALNLOQTAAQSAQKDLIES----- 442

QY 259 GSEFLQDFSETRYHTESLQNMKSQELIKEYLELEKCL-----SRM 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 ---LQOQITMWRKQYETLAKRYSSMR---EYLALKLKATQOAKSAKEALEKEM 495

QY 302 E-----DENNRILRESKRL-GGDDARVLELEL-----DLRAEN 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 ERDRHKNIELADIKERDRARYDLDRAGKGNKEDVERLERLMAQDLKDRSTGAD 555

QY 337 LQLL-----TENELHQOERAPLSKRGD 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 LSLLSKHNRELSELENAL-KMKORA-LDERGD 586
```

RESULT 13

B45344
Probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)
C:Species: suid herpesvirus 1
C:Date: 30-Sep-1993 #sequence, revision 30-Sep-1993 #text, change 16-Jul-1999
C:Accession: B45344
R:Virek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwytzer, M.
Virology 179, 365-377, 1990
A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented of
A:Reference number: A45344; MUID:91021039; PMID:2171211
A:Accession: B45344
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-1733 <VIC>
A:Cross-references: GB:M4651; NID:9334070; PIDN:AAA47471.1; PID:9334072
C:Superfamily: pseudorabies virus 1 nuclear antigen

Query Match 6.7%; Score 128; DB 1; Length 1733;
Best Local Similarity 22.4%; Pred. No. 5.4;
Matches 74; Conservative 27; Mismatches 94; Indels 136; Gaps 16;

```
8 EYHQPOTSCT-----GAAYQEE-----INPER-PEGAEER----- 39
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 DQGRPPPTTTTITKDPHODPLLPTLTLOEDPHLLRPTDRPSAKTNHHQDPPG 163

QY 40 VPEDDSRWQSRAPFOLGGRGP-----EGESLSEOPPLQ----- 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 164 GPPSTSHHHNQDPGGGPPPPRPPSTSSSSHQGPSTRRPPORRRMPPSPQKI 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 76 --TOACPESSCLR-----EGEKQONGDSSAGG-----PEPPAEVEPIPEA 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 SETRAGSENTAQTILFSHSEKTLFSHPMEGEGE-GDRGTAGGDEDRDDPPPPSPPPPP 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 116 ELLAQPCHDSASKLGAFAAGGEEMQOQOROLG---KKHRRPSSKKRHKPYKLTW 172
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 PLPPPPPPPPPPQ---PPPAGCS---ARRRRGGGPPGGRGRRGRRR----- 326

QY 173 EKKKFKDEKQSLRASRIAEEMFAKGQVAPYNTQFLMDHDOEPRLTGKLYSKRAAK 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 -----ACGTGAA-----AADAEEEDGD----- 344

QY 233 SDPTSDDPFMEEGEGEGSD--GMGDPGSE 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 GDEDEDEDRAKGEGRDQEGEGRAGGAGG 375
```

RESULT 14

T06310
hypothetical protein F11C18.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence, revision 30-Apr-1999 #text, change 22-Oct-1999
C:Accession: T06310
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysbaert, C.; Dasseville, R.; De Clerck, R.;
Eves, H.W.; Mayer, K.F.X.; Scheller, C.
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15589
A:Accession: T06310
A:Molecule type: DNA
A:Residues: 1-852 <BEV>
A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.80
A:Experimental source: cultivar Columbia; BAC clone F11C18
C:Genetics:
A:Gene: ATSP:F11C18.80
A:Map position: 4
A:Introns: 34/3; 102/3; 205/3; 282/3; 564/3; 602/3; 622/3; 639/3; 667/3

Query Match 6.7%; Score 127.5; DB 2; Length 852;
Best Local Similarity 23.2%; Pred. No. 2.8;
Matches 88; Conservative 51; Mismatches 145; Indels 95; Gaps 18;

```
1 MAEPFLSEYOHQPTNSCTGAAYQEEINPERPPGAEERYPEEDSRMSRAFPOLGGRPG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 439 MAEE-LPEQSVPKKTAQKKKESSTEEVKPSATATEE-VSEEPNTE---PQVTKKS 492

QY 61 PEGGSLSQPP-PLTOACPESSCLREGEKGQNGDSSAGDPPPAEVEP---TPEA 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 KKVASSSKTKPTVPSSKSTSEKTVARQSEKKVGSQMAQESTPKPEKKRPGKALDE 552

QY 116 ELLAQPCHDSF--ASKLGAFAAGGEEMQO-----OORLGKKHRRRP--SK 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 ESLHTSGDNEKPAVSSGKLASKKRAKQTVESPSNTRKRSLOQGAAGSELVGR 612

QY 160 KRRHW---KPYKLTWEE---KKF-----DEKQSLRASRIAEEMFAKQVAPYNTQ 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 IKVWPMQAVYKGVESYDAKKKHILVIYDDQDEI-----LYLKQKMSPLDESE 664

QY 208 FLMD-----DHQDEPDLKT---GLYSKRA-----AAKSDTSDDD----- 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 LSQDEAADQTOGEEDASTVGSAGSKAAATPASKSSKTSQDQKTAASKDSKDSK 724

QY 241 --FMEEGEGEGSDGSGSEFLQDFSETYRYHTESLQNMKSQELIKEYLELEKCL 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 EASSEESSEEPKTYGKSGSSKSKNDIS-----SVKSGK-----SKAS 765

QY 299 SRMEDENRRLRESKRLGG 317
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 SKKKEEPSKATVTSKSKSG 784
```

RESULT 15

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:09:56 ; Search time 24 Seconds
(without alignments)
620.417 Million cell updates/sec

Title: US-09-972-758a-2

Perfect score: 1910
Sequence: 1 MAEPPLSEYOHOPOTSNGT.....LTENELHROOERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	164.5	8.6	1224	1 DYNA_CHICK	P35458 gallus gall
2	139	7.3	756	1 BRS1_BOVIN	006002 bos taurus
3	137	7.2	536	1 GAG_MLVCH	P27460 cas-br-e mu
4	134.5	7.0	1181	1 NKX1_RAT	0942m6 rattus norv
5	134	7.0	568	1 INVO_RAT	P48998 rattus norv
6	133	7.0	538	1 GAG_MLVMO	P03332 moloney mur
7	133	7.0	1493	1 ERG6_HUMAN	003468 homo sapien
8	131	6.9	1898	1 TRHY_HUMAN	007283 homo sapien
9	130	6.8	466	1 CMGA_RAT	P10354 rattus norv
10	129.5	6.8	1816	1 AFG_HUMAN	P55196 homo sapien
11	128	6.7	1733	1 VNUA_PRYKA	P33485 pseudorale
12	128	6.7	2476	1 ATRX_MOUSE	P30427 rattus muscu
13	127.5	6.7	4687	1 PLEI_RAT	P30427 rattus muscu
14	126.5	6.6	1407	1 TRHY_RABIT	P3709 oryctolagus
15	125.5	6.6	1549	1 TRHY_SHEEP	P27953 ovis aries
16	125	6.5	2004	1 MOZ_HUMAN	092794 homo sapien
17	124	6.5	449	1 CMGA_BOVIN	P05059 bos taurus
18	123.5	6.5	677	1 SGL_MOUSE	P16014 mus musculu
19	122.5	6.4	4473	1 PLEI_CRIGR	09155 cricetus
20	122	6.4	723	1 SSRP_DROME	003344 drosophila
21	121	6.3	538	1 GAG_MSVMO	P03334 moloney mur
22	121	6.3	726	1 BRD3_HUMAN	015059 homo sapien
23	121	6.3	1089	1 Y553_HUMAN	094x13 homo sapien
24	121	6.3	1280	1 DYNA_RAT	P28023 rattus norv
25	120	6.3	536	1 GAG_MLVDE	P29168 rattus norv
26	120	6.3	699	1 NP14_HUMAN	019478 homo sapien
27	120	6.3	1781	1 AKAC_HUMAN	002952 homo sapien
28	119.5	6.3	556	1 HIR3_HUMAN	09bw71 homo sapien
29	118.5	6.2	472	1 TWIE_CAEEL	025525 caenorhabdi
30	118.5	6.2	764	1 UBFL_RAT	P25977 rattus norv
31	117.5	6.2	372	1 TOLA_HAEIN	P44678 haemophilus
32	117.5	6.2	513	1 DMP1_HUMAN	013316 homo sapien
33	117.5	6.2	2142	1 BAT2_HUMAN	P48634 homo sapien

34	117	6.1	568	1 RGS3_MOUSE	Q9d604 mus musculu
35	117	6.1	845	1 NFM_RAT	P12839 rattus norv
36	117	6.1	911	1 CAFI_MOUSE	Q94w10 mus musculu
37	116.5	6.1	503	1 DMP1_MOUSE	055188 mus musculu
38	116.5	6.1	1281	1 DYNA_MOUSE	008788 mus musculu
39	116	6.1	927	1 YOL3_CAEEL	002328 caenorhabdi
40	116	6.1	1729	1 TABP_HUMAN	Q9c0c2 homo sapien
41	116	6.1	1972	1 MYHB_RABIT	P35748 oryctolagus
42	115.5	6.0	793	1 CALD_HUMAN	005682 homo sapien
43	115.5	6.0	2472	1 NCR2_MOUSE	Q9wu42 mus musculu
44	115	6.0	1739	1 CHD2_HUMAN	014647 homo sapien
45	115	6.0	1745	1 ZOL_MOUSE	P39447 mus musculu

ALIGNMENTS

RESULT 1
DYNA_CHICK STANDARD: PRT: 1224 AA.
AC P35458:
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)
DE (p150-glued).
GN DCTNL.
OS Gallus gallus (chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Cleaveland D.W.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 172-1224 FROM N.A.
RC TISSUE=Embryonic brain;
RX MEDLINE=92098576; PubMed=1836789;
RA Gili S.R., Schroer T.A., Szliak I., Steuer E.R., Sheetz M.P.,
RT "Dynactin, a conserved, ubiquitously expressed component of an
RT activator of vesicle motility mediated by cytoplasmic dynein.";
RL J. Cell Biol. 115:1639-1650(1991).
CC -!- FUNCTION: DYNACTIN IS A MAJOR COMPONENT OF ACTIVATOR I. A 20S
CC POLYPEPTIDE COMPLEX THAT STIMULATES DYNEIN-MEDIATED VESICLE
CC TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
CC -!- SIMILARITY: BELONGS TO THE DYNACTIN 150 KDA SUBUNIT FAMILY.
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X62773; CAA44617.2; -.
CC PIR: A41642; A41642.
CC InterPro: IPR000938; CAP-Gly.
CC Pfam: PF01302; CAP-Gly. 1.
CC PROSITE: PS00845; CAP-Gly. 1.
CC PROSITE: PS50245; CAP-Gly. 2; 1.
CC Motor protein: Microtubules; Dynein; Coiled coil; Cytoskeleton;
CC Alternative splicing.
CC FT DOMAIN 49 91 CAP-GLY.
CC FT DOMAIN 205 540 COILED COIL (POTENTIAL).
CC

FT DOMAIN 936 1042 COILED COIL (POTENTIAL).
 FT DOMAIN 1081 1117 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1224 AA; 135562 MW; 03B7FEE68E7C01D7 CRC64;

Query Match
 Best Local Similarity 22.6%; Pred. No. 0.027; DB 1; Length 1224;
 Matches 83; Conservative 61; Mismatches 152; Indels 71; Gaps 14;

29 NPERPPGAEEYRVEDSRWQSRAPFOLGREGEGESLSEOPPILOTQACPESSCLREG 88
 106 SPETPESALAKVPRKHSRAAKGSKLRGAKP-----KTTARPKPTRTPTSPSS-----G 157
 QY 89 EKGQNDSSAGGDP--PPPAVEPTP-EAELLAPCHDSEASKLGAPAGGEEBWGOQO 145
 DB 158 TAGPSGASASGCGESSSEPTPAOTPLVAPVIPSPLTSPVAPVPSPTKEENLRSGV 217
 QY 146 ROLGK-----KHRRRPSKKRHWKPYKLTWEE-----KKFPKOSLRASRR-AMPAK 196
 DB 218 RDLERKLETLKTRKEDKAKLKLKELRYKIQLEOVEMSKQEQADJLORLKEAKKRAK 277
 QY 197 GQVAPRYNTTQFLMDHDQ-EEPDLTGLYSKRAAKSDTSDDDFMEGEGE----- 247
 DB 278 DALEAKERYMEMATPAIEMATLDKEMAEERASLQEV---DSLKEKVEYLTMDLEI 334
 QY 248 -----EDGSDGMDGSGEPLORDPSETYERYHTESLONMSKOEIik-----EYL 292
 DB 335 LKHEIEKSGSDA-----ASSYOVKOLEQONARLKEALVRMDLSASEKOEHV 382
 QY 293 ELEKCLSRMEDENNRRLRLESKRLGGDDARVRELLELDRLR-----AENIOLLLEN 343
 DB 383 KLQKMEKKNTELESRLROGREKL---QEEVKAETVDELKQVDAALGAEEMVETTLER 439
 QY 344 ELHROE 350
 DB 440 NIDLEEK 446

RESULT 2
 BFS1_BOVIN STANDARD: PRT; 756 AA.
 AC Q06002;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Filensin (Beaded filament structural protein 1).
 GN BFS1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID:9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Lens;
 RX MEDLINE=93260017; PubMed=8491777;
 RA Gounari F., Merdes A., Quinlan R., Hess J.F., Fitzgerald P.G.,
 RA Ouzounis C.A., Georgatos S.D.;
 RT "Bovine filensin possesses primary and secondary structure similarity
 RT to intermediate filament proteins.";
 RL J. Cell Biol. 121:847-853(1993).
 RN [2]
 RP REVISIONS, AND SEQUENCE FROM N.A.
 RC TISSUE=Lens;
 RA Hess J.F.;
 RA Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 CC -I- SUBUNIT: ASSOCIATES WITH BFS2.
 CC -I- SUBCELLULAR LOCATION: MEMBRANE- AND CYTOSKELETON-ASSOCIATED;
 CC -I- TISSUE SPECIFICITY: LENS.
 CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X72388; CAA51081.1; -;
 DR InterPro: IPR001664; IF.
 DR PROSITE: PS00226; IF; FALSE_NEG..
 KW Intermediate filament; Repeat; Membrane; Coiled coil; Cytoskeleton;
 KW Eye lens protein; Phosphorylation;
 FT DOMAIN 1 38 HEAD.
 FT DOMAIN 39 318 ROD.
 FT DOMAIN 319 755 TAIL.
 FT DOMAIN 39 73 COIL 1A.
 FT DOMAIN 74 82 LINKER 1.
 FT DOMAIN 83 182 COIL 1B.
 FT DOMAIN 183 199 LINKER 12.
 FT DOMAIN 200 318 COIL 2.
 FT DOMAIN 531 621 7 x 14 AA TANDEM REPEATS.
 FT REPEAT 531 544 1.
 FT REPEAT 545 551 2 (INCOMPLETE).
 FT REPEAT 552 565 3.
 FT REPEAT 566 579 4.
 FT REPEAT 580 593 5.
 FT REPEAT 594 607 6.
 FT REPEAT 608 621 7.
 FT MOD_RES 5 5 PHOSPHORYLATION (BY PKA) (POTENTIAL).
 SQ SEQUENCE 756 AA; 83001 MW; F86A18208A8E109 CRC64.

Query Match
 Best Local Similarity 7.3%; Score 139; DB 1; Length 756;
 Matches 88; Conservative 32; Mismatches 117; Indels 152; Gaps 16;

26 FELNPERPPGAEEYRVEDSRWQSRAPFOLGREGEGESLSEOPP-PILOTQ----- 78
 DB 415 EEASPTTQEGAPEDVP--DGCKISKAFELKMKIKKVKGRKEPPADLTTKGRYVWS 472
 QY 79 -----CPSSCLREGKGGQNDSSAGGDP-PPPAVEPTP-----EAELLAQ 120
 DB 473 GDGSFVDPGFCVFSVPAGGVVWSGDDS-----VPPDSGVPEPSPQOPEPLEGGQGPQ 527
 QY 121 PCHDSEASKLGAPAGG----- 587
 DB 528 EKEGGLKEEGPPREGKGPREGKGDVYKEGPPREGKGDVYKEGPPREGKGDVYKEG 587
 QY 143 QQQ-ROLGKKKRRRPSKKRHWKPYKLTWEEKKPEDEKOSLRASRLRAMPAGQVPA 201
 DB 588 PREGKGDVYKKEGPPREGKGLK-----EEGPIQEKD-----GOSPT 627
 QY 202 PYNTTQFLMDHDQEEPDLTGLYSKRAAAKSDTSDDDFMEG-----GEED- 249
 DB 628 PHPA-----DKGDEKNAKELKLGQKQ-----DDQKEGARGCPVAVAPGPGPS 672
 QY 250 -----GSDGMDGDSSEFLQR-----DSEYTERHTESLONMSKOEI 287
 DB 673 TPSPGQPOVILIGSGEGHARGSGRLARSPPKLAEKVEVWESIEKESPTIQTVEETAV 732
 QY 288 IKLEYLEKCLSRMEDENNRRLRLESKRLG 316
 DB 733 IVERTM-IER-----TKANKKILG 749

RESULT 3
 GAG_MLVCB STANDARD: PRT; 536 AA.
 AC P27460;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE GAG polypeptide [contains: Core protein p15; Inner coat protein p12;
 DE Core shell protein p30; Nucleoprotein p10].
 GN GAG.
 OS Cas-Br-E murine leukemia virus.

FT	DOMAIN	545	552	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	553	569	POTENTIAL.
FT	DOMAIN	570	989	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	990	1010	POTENTIAL.
FT	DOMAIN	1011	1017	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1018	1038	POTENTIAL.
FT	DOMAIN	1039	1053	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1054	1074	POTENTIAL.
FT	DOMAIN	1075	1092	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1093	1113	POTENTIAL.
FT	DOMAIN	1114	1121	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1122	1142	POTENTIAL.
FT	DOMAIN	1143	1150	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1151	1171	POTENTIAL.
FT	DOMAIN	1172	1181	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	461	501	ALPHA-1.
FT	REPEAT	1061	1092	ALPHA-2.
FT	DOMAIN	730	905	14 X APPROXIMATE TANDEM REPEATS.
FT	REPEAT	730	741	1.
FT	REPEAT	742	754	2.
FT	REPEAT	755	766	3.
FT	REPEAT	767	778	4.
FT	REPEAT	779	791	5.
FT	REPEAT	792	804	6.
FT	REPEAT	805	817	7.
FT	REPEAT	818	830	8.
FT	REPEAT	831	843	9.
FT	REPEAT	844	856	10.
FT	REPEAT	857	869	11.
FT	REPEAT	870	881	12.
FT	REPEAT	882	893	13.
FT	REPEAT	894	905	14.
FT	DOMAIN	952	974	POLY-GLU.
FT	MOD_RES	625	625	PHOSPHORYLATION (POTENTIAL).
FT	CARBOHYD	271	271	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	598	710	MISSING (IN ISOFORM 2).
FT	VARSPLIC	616	710	MISSING (IN ISOFORM 3).
FT	VARSPLIC	652	679	MISSING (IN ISOFORM 4).
Q	SEQUENCE	1181 AA:	129980 MM:	B065C1C119369AE CRC64;

	Query Match	7.0%; Score 134.5; DB 1; length 1181;
	Best Local Similarity	22.5%; pred. No. 1.2;
	Matches	67; Conservative 44; Mismatches 160; Indels 27; Gaps 7.
OY	19 TGAAYOEBLMPERPCGAEEKVPREDTSW---- <td>73</td>	73
Dd	692 TPAPARDPKGOEEDPCCOEVDVAENHGDMTGTGEGERETEAEGKKDEGETEARKEBDG	751
OY	74 LOTACPESSCLRGEEKQNDDSDSAGDPPRAPEVETTPAELLAOEC-HDSFAKIG-	131
Dd	752 QEETETKKGKCKOCSTPESBECKDOEGTEAEKADHEGETEAEGKVEYHNGETEACT	811
OY	132 -----APAAGEEMGGQQOLKKKHRRRPSKKKNMKRYUKLTWEKKKFKDEKSL	184
Dd	812 EDEOCEGETEAEGKEVEOGETEAEGKEVEHEVEUTEAEKKTENNHEGETEAEGKEADHNEGET	871
OY	185 RAS--KIRAEFMFGQPVARYNTPTQLMDND-----QEPRDKTLTYSKRAAKADD	235
Dd	872 EAEGNVTHOGTEAEKVEHEGTEAGCKDHDEGOSQTQADTEKCD--EGEAANAED	929
OY	236 TSDDDFMEEGEGEEDSGSDMGDSSEFLORDPFSETERYKHNSLOMMSKOELIKLEYLE	293
Dd	930 QCETAQGEKAGDGCGGSD--GCDSDEBEDDEDDEEEEEEEBEEBEESEFPUSLEMP	965
RESULT 5		
INVO_RAT	ID_INVO_RAT	STANDARD: PRT: 568 AA.
AC	P48998:	
DT	01-FEB-1996 (Rel. 33, Created)	
DT	01-FEB-1996 (Rel. 33, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DT	Involucrin.	

Q	SEQUENCE	568 AA: 67021 MW: E2B07740FE779345 CRC64;
K	Keratinization; Repeat.	
DR	PROSITE, PS00795; INVOLUCRIN; 1.	
DR	InterPro: IPR002360; Involutrin.	
EMBL; L28818; AAA41445.1; -		
CC	or send an email to license@lsb.slb.ch).	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.lsb.slb.ch/announce/	
CC	use by non-profit institutions as long as its content is in no way	
CC	between the European Bioinformatics Institute. There are no restrictions on its	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	- - SIMILARITY: BELONGS TO THE INVOLUCRIN FAMILY.	
CC	molecules.	
CC	are cross-linked to keratins, desmoplakin and to inter involucrin	
CC	- PFM: Substrate of transglutaminase. Specific glutamines or lysines	
CC	stratified squamous epithelia.	
CC	- TISSUE SPECIFICITY: Keratinocytes of epidermis and other	
CC	of the cornified envelope.	
CC	- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding	
CC	stratified squamous epithelia.	
CC	- FUNCTION: Part of the insoluble cornified cell envelope (CE) of	
CC	MOL. Biol. Evol. 10:1136-1149(1993).	
RT	repeats.";	
RA	Green H.;	
RA	"The involucrin genes of the mouse and the rat: study of their shared	
RA	Djian P., Phillips M., Easley K., Huang E., Simon M., Rice R.H.,	
RX	MEDLINE=94104476; PubMed=8277848.	
RC	SHRAIN-Sprague-Dawley;	
RP	SEQUENCE FROM N.A.	
RN	[1]	
OX	NCBI_TaxID=101116;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OS	Rattus norvegicus (Rat).	
IVL.		

```

Query Match      7.0% ; Score 134 ; DB 1 ; Length 568 ;
Best Local Similarity 20.9% ; Pred. No. 0.61 ;
Matches 93 ; Conservative 66 ; Mismatches 169 ; Indels 118 ; Gaps 18 ;

QY      3 EPELSEYONOROTSNSTGAAYOE-----LNPERRGAEERYEED 44
      1 : : : : : : : : : : : : : : : : : : : : : : : :
Db      127 ENHLDHNRDQRESOSGGLSGDQDVLARDELHMGNOKEKIDRELLRIGDQKTBEEQ 186
      1 : : : : : : : : : : : : : : : : : : : : : : : :

QY      45 -----SRMGRARPOLSGRGPR-----GEVSLESORPLDTQAS-----PESSCLR 86
      1 : : : : : : : : : : : : : : : : : : : : : : : :
Db      187 ELLHGEKQOKLILVERHNRQERQEBELNNGOKQKQDQROEYOLATQVONOKONERE-ICLR 245
      1 : : : : : : : : : : : : : : : : : : : : : : : :

QY      87 EGEKGGNGDDSSAGDFRRPAVEYR-----TREAELLARQCHNSEASKIGAR 133
      1 : : : : : : : : : : : : : : : : : : : : : : : :
Db      246 KQDQDQSHNEHELNLGKQDQDQSHNEHELNLGKQDQDQSHNEHELNLGKQDQDSECEL--- 302
      1 : : : : : : : : : : : : : : : : : : : : : : : :

QY      134 AAGSEEWGSGQDQDQ-----LQKKKNR-----RPSKKKNRYKYLKTEKKKF- 178
      1 : : : : : : : : : : : : : : : : : : : : : : : :
Db      303 NLSGQDQDQDQDQDQSECELNLGKQDQDQDQDQSELDLGKQDQDQDQDQDQDQDQDQDQK 362
      1 : : : : : : : : : : : : : : : : : : : : : : : :

QY      179 -----DEKQSLRASRIKAEI-----FAKQDVARUYNITQFLMIDNDQER 218
      1 : : : : : : : : : : : : : : : : : : : : : : : :
Db      363 DPELNLGKQDQDQDQSELDQVERKQDQSECELNLGKQDQDQDQDQDQDQDQDQDQK 422
      1 : : : : : : : : : : : : : : : : : : : : : : : :

QY      219 DLKTLGYSKRAAKSDQDQDQDQMEEGSEEDGSDGSGSGSEFLQDQDQETERYNTES 278
      1 : : : : : : : : : : : : : : : : : : : : : : : :
Db      423 EHLHG-----KQDSEHMDTEPQEKQDQSELYERLNLGKQDQDQDQDQDQDQDQDQDQK 472
      1 : : : : : : : : : : : : : : : : : : : : : : : :

QY      279 LQMSKQOEILKEYLELQKSLRMEDENNRL--RLSKSLGSGDQDQAVELDELQDRLAEI 336
      1 : : : : : : : : : : : : : : : : : : : : : : : :
Db      473 LMDLQKQKASQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQK 524
      1 : : : : : : : : : : : : : : : : : : : : : : : :

QY      337 L-----QLLTENELHROQERAPL 354
      1 : : : : : : : : : : : : : : : : : : : : : : : :

```

DB 525 LTOKEKQIKQLYPSTD--RVQETQPI 548

RESULT 6

GAG_MLVMO STANDARD; PRT; 538 AA.

ID GAG_MLVMO

AC P03332;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE GAG polypeptide [containing: Core protein P15; Inner coat protein P12; Core shell protein P30; Nucleoprotein P10].

GN GAG.

OS Moloney murine leukemia virus.

OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.

OX NCBI_TaxID=11801;

RP SEQUENCE FROM N.A. (CLONE PHLV-1).

RX MEDLINE=82035843; PubMed=6169994;

RA Shindick T.M., Lerner R.A., Sutcliffe J.G.;

RT "Nucleotide sequence of Moloney murine leukaemia virus.";

RL Nature 293:543-548(1981).

RP SEQUENCE OF 2-31, AND MYRISTOYLATION.

RX MEDLINE=83169654; PubMed=6340098;

RA Henderson L.E., Kruttsch H.C., Oroszlan S.;

RT "Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational proteins modification.";

RL Proc. Natl. Acad. Sci. U.S.A. 80:339-343(1983).

RN [3]

RP SEQUENCE OF 479-529.

RX MEDLINE=81264245; PubMed=6267042;

RA Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W., Oroszlan S.;

RT "Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses.";

RL J. Biol. Chem. 256:8400-8406(1981).

CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POLYPROTEIN.

CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: J02255; AAB59942.1; -

DR PIR: A03930; FOMVIM.

DR InterPro: IPR000840; Gag_MA.

DR InterPro: IPR002079; Gag_P12.

DR InterPro: IPR003036; Gag_P30.

DR InterPro: IPR001878; ZnF_CCHC.

DR Pfam: PF00098; zf-CCHC; 1.

DR Pfam: PF01140; Gag_MA; 1.

DR Pfam: PF01141; Gag_P12; 1.

DR Pfam: PF02093; Gag_P30; 1.

DR SMART: SM00343; ZnF_C2HC; 1.

DR PROSITE: PS50158; ZF_CCHC; 1.

DR Core protein: Coat protein; Nucleoprotein; Polypeptide; Myristate; Zinc-finger.

KW Core protein: Coat protein; Nucleoprotein; Polypeptide; Myristate; Zinc-finger.

RN

FT CHAIN 2 131 CORE PROTEIN P15.

FT CHAIN 132 215 INNER COAT PROTEIN P12.

FT CHAIN 216 478 CORE SHELL PROTEIN P30.

FT CHAIN 479 534 NUCLEOPROTEIN P10.

FT ZN_FING 502 519 CCHC-TYPE.

FT LIPID 2 2 MYRISTATE.

SO SEQUENCE 538 AA; 60858 MW; 8A7652439B464495 CRC64;

Query Match 7.0%; Score 133; DB 1; Length 538;

Best Local Similarity 20.9%; Pred. No. 0.66;

Matches 89; Conservative 40; Mismatches 142; Indels 154; Gaps 19;

QY 30 PERPGA-----EEVPEEDSRMOSRAFPOLGGRGPE---GECS-----LESDPPL 74

DB 107 PLPLPSAPSLPLEPPSTPPRSSLYPALT-PSLGAKPKQVLSGCGPLIDLTEDPP 165

QY 75 QTOQACPESSCLREGEGQNGDDSSAGDPPPAEVEPTPEAEEL-----AQPCDSEASKL 130

DB 166 RDRRPPSD--RGC-----NGEATPAGCAP-----DPSPMASRLRCRRREPVAIDTTSSQA 214

QY 131 GAPAGGEEBEGQOQOLGKKHRRRPSKKRRHWKPYKLTWEKKKPKDEKOSLRASRLR 190

DB 215 FPLRAGGN-----CQIQYWPSSSDLYNMK-----NNNPSFSDPKLTALIE 257

QY 191 AEMFAKGPVAPYNTTQPLMDHDHGEPPDKTG-----LYSKRAAKSND 235

DB 258 SVLI-----THQPTWDDCQQLLGLTLTGEKQRYVLEARKKAVRGDGRPTOLPNE 307

QY 236 -----TSDDPFMEGGE-----DGSDMGGDGSE 261

DB 308 VDAPFLERPDMDYTTQACGNHLVHYRQILLACLOAGRSPTNLAKVKGITGCPNESP 367

QY 262 FLDRPFSEYERY-----HRESLOMSKOELIKE 290

DB 368 FLER-LKEAVRRYTPYDPEDPGQETNVMSFTWQAPDGRKLERLEDLKNKTLGDLVR- 425

QY 291 YLELEKCSMEDENNRLIESKRLGGDARV-RELELELDRLRAENLQLTITNELHROO 349

DB 426 --EAERIFNKRREPEER-----EERIRRETEKERRRTEDEKKEKRRRRHR 472

QY 350 ERAVL 354

DB 473 EMSKL 477

RESULT 7

ERC6_HUMAN STANDARD; PRT; 1493 AA.

ID ERC6_HUMAN

AC 003468;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Excision repair protein ERCC-6 (Cockayne syndrome protein CSB).

GN ERCC6 OR CSB.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93092214; PubMed=1339317;

RA Troelstra C., van Gool A., de Wit J., Vermeulen W., Bootsma D., Hoeijmakers J.H.J.;

RT "ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes.";

RL Cell 71:939-953(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93181229; PubMed=8382798;

RA Troelstra C., Hesen V., Bootsma D., Hoeijmakers J.H.J.;

RT "Structure and expression of the excision repair gene ERCC6, involved in the human disorder Cockayne's syndrome group B.";

RL Nucleic Acids Res. 21:419-426(1993).

RN [3]

RP REVIEW ON VARIANTS CSB.

RX MEDLINE=99374920; PubMed=10447254;

RA Cleaver J.E., Thompson L.H., Richardson A.S., States J.C.;

RT "A summary of mutations in the UV-sensitive disorders: Xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy.";

RL Hum. Mutat. 14:9-22(1995).

RN [4]

RP VARIANTS CSB, AND VARIANTS.

RX MEDLINE=98107940; PubMed=9443879;

RA Mallery D.L., Tanganeli B., Colella S., Steingrimsdottir H.,
 RA van Gool A.J., Troelstra C., Stefanini M., Lehmann A.R.;
 RT "Molecular analysis of mutations in the CSB (ERCC6) gene in patients
 RT with Cockayne syndrome.";
 RL Am. J. Hum. Genet. 62:77-85(1998).
 CC -1- FUNCTION: IS INVOLVED IN THE PREFERENTIAL REPAIR OF ACTIVE GENES.
 CC PRESUMED DNA OR RNA UNWINDING FUNCTION. CORRECTS THE UV SURVIVAL
 CC AND RNA SYNTHESIS AFTER UV EXPOSURE OF COCKAYNE'S SYNDROME
 CC COMPLEMENTATION GROUP B.
 CC -1- SUBUNIT: INTERACTS WITH THE CSB PROTEIN AND A SUBUNIT OF RNA
 CC POLYMERASE II TFIIH.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- DISEASE: Defects in ERCC6 are the cause of Cockayne's syndrome
 CC type B (CSB). CSB is a disease which is characterized by dwarfism,
 CC precociously senile appearance, pigmentary retinal degeneration,
 CC optic atrophy, deafness, sensitivity to sunlight, and mental
 CC retardation.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC -1- DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chromocancer/genes/CSBID302.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L04791; AAA52397.1; -;
 DR PIR: A44224; A44224;
 DR TRANSFAC: T04550; -;
 DR Genew: HGNC:3438; ERCC6.
 DR MIM: 133540; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N.
 DR Pfam: PF00176; SNF2_N.1.
 DR Pfam: PF00271; helicase_C.1.
 DR SMART: SM00487; DEXDC.1.
 DR SMART: SM00490; HELICC.1.
 DR Helicase; DNA repair; ATP-binding; DNA-binding; Nuclear protein;
 DR Transcription regulation; Cockayne's syndrome; Disease mutation;
 KW Polymorphism.
 KM Polymorphism.
 FT DOMAIN 356 394 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 442 446 GLY-RICH.
 FT DOMAIN 466 481 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 1038 1055 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT NF-BIND 532 539 ATP (POTENTIAL).
 FT SITE 646 649 DEGH BOX.
 FT VARIANT 255 255 K -> T.
 FT VARIANT 399 399 /FTID=VAR_001216.
 FT VARIANT 670 670 G -> D.
 FT VARIANT 670 670 /FTID=VAR_001217.
 FT VARIANT 851 851 R -> W (IN CSB).
 FT VARIANT 851 851 /FTID=VAR_001218.
 FT VARIANT 957 957 W -> R (IN CSB).
 FT VARIANT 957 957 /FTID=VAR_001219.
 FT VARIANT 1042 1042 V -> G (IN CSB).
 FT VARIANT 1042 1042 /FTID=VAR_001220.
 FT VARIANT 1095 1095 P -> L (IN CSB).
 FT VARIANT 1095 1095 /FTID=VAR_001221.
 FT VARIANT 1097 1097 P -> R (IN CSB).
 FT VARIANT 1097 1097 /FTID=VAR_001222.
 FT VARIANT 1213 1213 M -> V.
 FT VARIANT 1213 1213 /FTID=VAR_001223.
 FT VARIANT 1413 1413 R -> G (IN CSB).
 FT VARIANT 1413 1413 /FTID=VAR_001224.
 FT VARIANT 1413 1413 O -> R.
 FT SEQUENCE 1493 AA: 168415 MM: 285257E2AEC071AC CRC64;
 Query Match 7.0%; Score 133; DB 1; Length 1493;

Best Local Similarity 23.7%; Pred. No. 1.9;
 Matches 82; Conservative 42; Mismatches 118; Indels 104; Gaps 17;
 QY 35 GAERYP-----EEDSRMOSRAFPOLGCRPRGEGSGLESOPPPLOTQACPPSSCLRGG 88
 Db 201 GAEVKTIELDHLASLEEDA-----EPGP-----SLGSLMIPVOETAWME--LIRIG 243
 QY 89 EKQCONDSSAGDFPPPAVEPPTP-----EA-----ELLAQPCDHSEASK 129
 Db 244 QM-----TPFGQITQOKQEKRRKIMLNASFEKYLADQALTSERKKQGNKRAAK 297
 QY 130 LGAPA-----AGCEEMGQOOROLGKKHRRSRKRRHKKPYKLTWEKKKFEKQSL 184
 Db 298 APAPVPPAPVQKNKPNKARVLSKKEE-----LKKHKIKQKRALQFOGKVGLPKAR 352
 QY 185 RA-----SRIRAEPAFGQAPAPYNTQFLMDHDQ-----EPDL-----KRG 223
 Db 353 RPWESDMRPAEEDSGEE-SEYPTPEEEEDDEVEGAEADLSGDGTDYELKPLPKG 411
 QY 224 LYSKRAAKSDOTSDDDFMEEGEE-DGSGDGGGSGSEFLQDFSETERYHTESLQNM 282
 Db 412 KRQKKVPVQE---IDDDFFSSGEELAMASVGGGGGGRVGRYRDDGDEYKQRLRW 468
 QY 283 SKQELIKEYLELEKCLSMEDENNRLRLSKRLGDDARVLELE 328
 Db 469 NK-----LRLQDKERUKLE-----DDSESDAEFD 494
 RESULT 8
 TRHY_HUMAN
 ID TRHY_HUMAN STANDARD; PRT; 1898 AA.
 AC Q07283;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trichohyalin.
 GN THH OR TRHY OR THL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93280194; PubMed=7685034;
 RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
 RA Steinhert P.M.;
 RT "The structure of human trichohyalin. Potential multiple roles as a
 RT functional EF-hand-like calcium-binding protein, a cornified cell
 RT envelope precursor, and an intermediate filament-associated (cross-
 RT linking) protein.";
 RT J. Biol. Chem. 268:12164-12176(1993).
 RN [2]
 RP SEQUENCE OF 1731-1898 FROM N.A. AND CHARACTERIZATION.
 RX MEDLINE=93315897; PubMed=7686933;
 RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinhert P.M.;
 RA "Trichohyalin: a structural protein of hair, tongue, nail, and
 RA epidermis.";
 RL J. Invest. Dermatol. 101:655-715(1993).
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 CC DIFFERENTIATION.
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA AND IN
 CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
 CC THE EPIDERMIS.


```

CC -1- SIMILARITY: BELONGS TO THE CHROMOGRAININ / SECRETINOGRANIN PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X06832; CA23988.1; -
DR PIR: A23996; A23996.
DR PIR: S00291; S00291.
DR PIR: S02543; S02543.
DR HSP: P05059; ICFK.
DR InterPro: IPR001819; Chromogranin_AR.
DR InterPro: IPR001990; Granin.
DR Pfam: PF01271; Granin.1.
DR PRINTS: PR00659; CHROMOGRAININ.
DR PROSITE: PS00422; GRANINS.1; 1.
DR PROSITE: PS00423; GRANINS.2; 1.
KW Signal: Amidation; Glycoprotein; Calcium-binding.
FT SIGNAL 1 18
FT CHAIN 19 466 CHROMOGRAININ A.
FT PEPTIDE 19 146 BETA-GRANIN.
FT PEPTIDE 281 332 PANCREASTATIN (POTENTIAL).
FT PEPTIDE 361 374 WE-14.
FT DOMAIN 92 110 POLY-GLN.
FT DOMAIN 231 249 POLY-GLU.
FT DOMAIN 344 350 POLY-GLU.
FT MOD_RES 332 332 AMIDATION (G-333 PROVIDE AMIDE GROUP)
FT (PROBABLY).
FT DISULFID 35 56 BY SIMILARITY.
SQ SEQUENCE 466 AA; 52024 MW; 05D135F8A657C48C CRC64;
Query Match 6.88; Score 130; DB 1; Length 466;
Best Local Similarity 24.08; Pred. No. 0.84;
Matches 87; Conservative 55; Mismatches 159; Indels 62; Gaps 17;
QY 3 EPTLSEVQHQPOTSNCGAAYOEELNPERPPGAEEVPEDESRWQSRAPFOLGR---P 59
DB 64 ERYLSIRHO-----NLKLELDLALOGAKERAQOQOQOQOQOQOQOQOQOHS 113
QY 60 GPEEGEGS--DESQPPPLQ---TQACPSSCLREGEKGQGDSSAG--DFP--PPAIVE 110
DB 114 SPEDELSEVENOSPAKKHGDASSEASKDTVEKRDSKQGDADFEGTTEGPRQAFPE 173
QY 111 PTREAEIL--AQPCHESEAKLAPAAAGEEEMGQOQOLGKKKRRRPS--KKRRHWK 166
DB 174 PKOESSMNGNSQSGEDTANNQSPSLPSQEHGIPQTTGSE---RGPAAOQOARRAKO 230
QY 167 YKLTWEKKKPEKOSLRASRIAREMFAGQVAPYNT-----TQFLMDHDDQEPDL 220
DB 231 EEKEEEEEEKEEEKEEKAIRKRGKPEVPTAASSHFYSGYKTIQKDDGQSSQSA 290
QY 221 KTLGYSKRAAKSDDTSDDFMEEGEGEDGSDGM-----GGDSEFLORDSESY 271
DB 291 VNG--KTGASEAVPSGKGLSHSQEEDGEEFAMGPPGGLRPGGKGP--LRKQDEE 346
QY 272 ERYHTESLQWMSQOLIKETLELEKCLSMEDENNRLRLSKRLGGDADARLELELDR 331
DB 347 E-----EEERLSREW--EDKRWSRMDQLAKELTAE--KRLEGEDDPDRSMKLSF-R 392
QY 332 LRA 334
DB 393 ARA 395
RESULT 10
AF6_HUMAN STANDARD; PRT; 1816 AA.
AC P55196; 075087; 075088; 075089; 09NU92;

```

KW Chromosomal translocation; Proto-oncogene; Alternative splicing.
FT DOMAIN 36 206 RAS-INTERACTING.
FT DOMAIN 425 491 FHA.
FT DOMAIN 804 910 DILUTE.
FT DOMAIN 991 1077 PDZ.
FT DOMAIN 162 174 GLU/LYS-RICH.
FT DOMAIN 1349 1356 POLY-PRO.
FT DOMAIN 1371 1376 POLY-PRO.
FT DOMAIN 1561 1571 ASP/GLU-RICH (ACIDIC).
SITE 26 26 ML FUSION POINT (IN AN ACUTE MYELOID
LEUKEMIA PATIENT).
FT VARSPLIC 1588 1611 LÖDERERQOOLEMRREADRA -> VKGVLMLCPSTV
FT VARSPLIC 1612 1816 PLIASACEPMG (IN ISOFORM 1).
FT VARSPLIC 1666 1743 MISSING (IN ISOFORM 1).
FT LCRPPLPDYPPSPAPGAPPPPPPOPMASLYLKTQVLSPD
FT PSTAKFYANNEEEEDCCSLAGODKSSMPKSHGDL ->
FT PNSTPGTGAAGVGAHADACROKERSKSDADSGSSCAPE
FT NUTKERORLFSGQDVSNKVKASRKLTLELELNTRK (IN
ISOFORM 3).
FT VARSPLIC 1744 1816 MISSING (IN ISOFORM 3).
FT CONFLICT 373 373 G -> V (IN REF. 1).
FT CONFLICT 391 391 P -> PGRNHFAVYNTHTYE (IN REF. 3).
FT CONFLICT 744 744 D -> DSHED (IN REF. 3).
FT CONFLICT 1031 1031 D -> DV (IN REF. 1).
FT CONFLICT 1408 1408 R -> P (IN REF. 1).
SQ SEQUENCE 1816 AA; 205604 MW; EB1FE7F04879CE8F CRC64;

Query Match 6.8%; Score 129.5; DB 1; Length 1816;
Best Local Similarity 20.3%; Pred. No. 3.7;
Matches 78; Conservative 51; Mismatches 144; Indels 111; Gaps 14;

QY 14 QTSNCTGAAYVOEELN--PERPPGAEEVPEEDSRMOS--RAFPOLGGRPEGECSLESQ 70
DB 1291 QSSLSIDSTSSQEHNLNSSKSVTPASTLTGSGPRGKMTPTAIPATPAVAVSQPIRTDLP 1350
QY 71 PPPIQTOACPESSCLREGCKNGDDSDSAGDPPPEVEPTPAELLAOPCHDSEASKL 130
DB 1351 PPPPVHYA-----GDFGMSMDLPLP---PPP-----SANOI 1380
QY 131 GAPPAAGGEEMGQOOROLGKKKKHRRSPKKRHKPYUKLTWE--EKKKFKDEKOSLRASRI 189
DB 1381 GLPFA-----OVAALERRKREHOMWYERKEKARLEEEERKRRERLGGOMRT 1429
QY 190 RAEMFANGQVPARYNTQF-----LMDHDQEEP-----DLKTGLYSKR 228
DB 1430 QS---LNPAPFSPLTAAQMKPEKPTLQRPQETVIRELQPOQRTTERDLQYITYTSKE 1486
QY 229 AAASDSDTSDDEMEEGEGEDGSGDGSGSEFLQRPDFTETRYHTESLQNNSKOELI 288
DB 1487 ELSSGDSLSPDPW-----KRDAKKELEKQOQMHTVDM-----LS 1520
QY 289 KEYLELEKCLSRMEDENNRLR-----LESKRLLGGDARVRELE-----LDRL 332
DB 1521 KEIPELOSKPRPSAEESRLKLMLEWFOFKRLOESKOKDEDEEEDDDVDTMLIMORL 1580
QY 333 RAENMLLTENELHRQOERAPLSK 356
DB 1581 EAFERRARLODEERRRQOOLEMRK 1604
RESULT 11
VNUA_VNUA STANDARD: PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable nuclear antigen.
OS Pseudotabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stages; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirinae;
OX NCBI_TaxID=33703;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; Pubmed=2171211;
RA Vilek C., Kozmik Z., Paces V., Schirm S., Schwyzer M.
RT "Pseudotabies virus immediate-early gene overlaps with an oppositely
RT oriented open reading frame: characterization of their promoter and
RT enhancer regions.";
RL Virology 179:365-377(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M34651; AAA47471.1; -;
DR PIR; B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CDBBEA75BB5E2 CRC64;

Query Match 6.7%; Score 128; DB 1; Length 1733;
Best Local Similarity 22.4%; Pred. No. 4.3;
Matches 74; Conservative 27; Mismatches 94; Indels 136; Gaps 16;

QY 8 EYOHQPTSNCT-----GAAAYOEE-----LNPER-PPGAEEER----- 39
DB 104 DHQHRPPRTTTTITKQPHQBDLRLPYTLQDEEDHLLRPTYPDPSSAKTHNHQDPGC 163
QY 40 VPEEDSRMOSRAFPOLGGRPP-----EGEGSLESQPPPLQ----- 75
DB 164 GPRSTSHNNHQDPGCGPRSPRPSTSSSHSQGPRSTRPPQPPPPPPPPPPQK1 223
QY 76 --TQACESSCLR-----EGEKQNGDDSDSAG--DPPPAVEEPTPEA 115
DB 224 SETRAGSENNYAQTLSHSEKNLFSHPMEGEGE-GDRGTAGGEGEDRDPPPPPPPPPP 282
QY 116 ELAQPCDHSEASKLGAAPAGGEEMGQOOROLA---KKHRRRSPKKRHKPYUKLTW 172
DB 283 PLPPPPPPPPPPQ---PPPAGCS---ARRRRGGGPPGRCGRKRGKRRR----- 326
QY 173 EKKKFKDEKOSLRASRIAEEMFAKQVPARYNTQFLMDHDQEEPDLKTGLYSKRAAK 232
DB 327 -----AGCTEAA-----AADAEEFEDGD----- 344
QY 233 SDDTSDDDFMEEGEGEDGSGD--GMGDCGE 261
DB 345 GDEDEDEDRAEGEGEDGEGPRGAGGAGE 375

RESULT 12
ATRX_MOUSE
ID ATRX_MOUSE STANDARD: PRT; 2476 AA.
AC Q61687;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked nuclear protein)
DE (heterochromatin protein 2) (HPI alpha-interacting protein) (HPI-BP38
DE protein).
GN ATRX OR XNP OR HPIBP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.

RX MEDLINE=98213653; PubMed=9545503;
 RA Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;
 RT "Comparison of the human and murine ATRX gene identifies highly
 RT conserved, functionally important domains";
 RL Mamm. Genome 9:400-403(1998).
 RN [2]
 RP SEQUENCE OF 325-1176 FROM N.A.
 RX MEDLINE=97133299; PubMed=8978696;
 RA le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
 RA Jeanmougin F., Losson R., Chambon P.;
 RT "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
 RT control of transcription by nuclear receptors";
 RL EMBO J. 15:6701-6715(1996).
 RN [3]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE=20040663; PubMed=10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
 CC GENE EXPRESSION BY AFFECTING CHROMATIN.
 CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 CC INTERACTING WITH HPI.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isp-slb.ch/announce/>
 CC or send an email to license@isp-slb.ch).
 CC -----
 DR EMBL: AF026032; AAC08741.1; .
 DR EMBL: X99643; CA67962.1; .
 DR MGD: MGI:103067; Xnp.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N.
 DR InterPro: IPR001841; ZnF_Ring.
 DR Pfam: PF00176; SNF2_N; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00184; RING; 1.
 KM DNA repair: Nuclear protein; DNA-binding; Helicase; ATP-binding;
 KM Zinc-finger.
 FT ZN_FING 219 267 PHD-TYPE.
 FT NP_BIND 1579 1586 ATP (POTENTIAL).
 FT SITE 1704 1707 DEGH BOX.
 FT DOMAIN 319 322 POLY-SER.
 FT DOMAIN 735 738 POLY-SER.
 FT DOMAIN 1001 1004 POLY-SER.
 FT DOMAIN 1130 1135 POLY-SER.
 FT DOMAIN 1182 1185 POLY-SER.
 FT DOMAIN 1238 1245 POLY-ASP.
 FT DOMAIN 1484 1487 POLY-GLU.
 FT DOMAIN 1924 1931 POLY-SER.
 FT DOMAIN 2205 2208 POLY-LYS.
 FT DOMAIN 2245 2248 POLY-GLU.
 FT DOMAIN 2403 2408 POLY-GLN.
 SO SEQUENCE 2476 AA; 278601 MW; 90A42B790FC4FF4C CRC64;

Query Match 6.7%; Score 128; DB 1; Length 2476;
 Best Local Similarity 19.7%; Pred. No. 6.2;
 Matches 78; Conservative 52; Mismatches 130; Indels 136; Gaps 13;
 QY 43 EDSRQSRAPFQOLGPEPEGE---GSLSEQPPPLQTOACPESSCLREGEKGQNGDDSSA 99
 DB 1142 EDNRKQKQRTSAKKKTGNTKEKRNLSLRATPKROYDITSSSPDGDDNSAGSESSD 1201
 QY 100 GGDPPP-----PAVEP-----TPE-----AELLA 119
 DB 1202 EOKIKPTVENLVLSHTGFCQSSGDALSKSVPAIVDDDDNDPBNRIAKMLLEETKA 1261
 QY 120 QPCHDSEASKLGAPAGGEEMGQOQ-----RQAGKKRRRPRSKK- 161
 DB 1262 NLSSDEGSSDDEPDGCKKRTGKQSESPADDELREGQLAVQNVSESSDSESKRP 1321
 QY 162 --RHKKPYKLTWEKKKFKDEKQSLRASRIAEKAKQPAVPYNTTQFLMDHDQDEPD 219
 DB 1322 RYRHLRLHKKLTLSGSEGEK---PKRPREKAKGR----- 1356
 QY 220 LKTGLYSKRAAKSDTSDDFMEGEGEDGSGMGSGSEFLQRPSETERYHTESL 279
 DB 1357 -----NRKVSSEDSDDTFOESGVSEVSE-----SEDRQRPRTSAKKALEEN 1402
 QY 280 QMSKQELIKYLELEKCLSR-----MEDENNRLRESKRL 315
 DB 1403 QRSYKQKKRRRIKQVQEDSSSEKSHSEDEKKEGDEDEDEDEDEDEND---DSKSP 1458
 QY 316 GGDARVLELELDRLAEMLQLTTELHROGR 351
 DB 1459 GKGRKIRKI-LKDDKLKTE-----TONALKEEER 1488
 RESULT 13
 ID PLE1_RAT STANDARD; PRT; 4687 AA.
 AC P30427; 008879; 008880; 008881;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Plectin 1 (PLTN) (PCN).
 GN Plect1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Glial tumor;
 RX MEDLINE=91268156; PubMed=2050743;
 RA Wiche G., Becker B., Lubert K., Weitzer G., Castanon M.J.,
 RA Hauptmann R., Stralowa C., Stewart M.;
 RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
 RT chain with a three-domain structure based on a central alpha-helical
 RT coiled coil";
 RL J. Cell Biol. 114:83-99(1991).
 RN [2]
 RP REVISIONS.
 RC TISSUE=Glial tumor;
 RX MEDLINE=96210632; PubMed=8633055;
 RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
 RT "Human plectin: organization of the gene, sequence analysis, and
 RT chromosome localization (8q24).";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
 RN [3]
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4), AND TISSUE SPECIFICITY.
 RC TISSUE=Glial tumor;
 RX MEDLINE=97321050; PubMed=9177781;
 RA Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
 RA Wiche G.;
 RT "Plectin transcript diversity: identification and tissue distribution
 RT of variants with distinct first coding exons and rodless isoforms";
 RL Genomics 42:115-125(1997).

CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
 CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
 CC HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
 CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
 CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
 CC SKELETAL MUSCLE AND LOWEST IN THYMUS.
 CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
 CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
 CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
 CC -1- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
 CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 PLECTIN REPEATS.
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
 CC -1- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC -----
 DR EMBL; X59601; CAA42169.1; -
 DR EMBL; U96274; AAC53209.1; -
 DR EMBL; U96275; AAC53210.1; -
 DR EMBL; U96276; AAC53211.1; -
 DR PIR; A39638; A39638.
 DR PIR; S21876; S21876.
 DR HSSP; Q01082; 1BKR.
 DR Interpro: IPR001589; Actbind_actnin.
 DR Interpro: IPR001715; Calponin-like.
 DR Interpro: IPR001101; Plectin_repeat.
 DR Interpro: IPR005326; S10_Plectin.
 DR Interpro: IPR002017; Spectrin.
 DR Pfam; PF00307; CH; 2.
 DR Pfam; PF00681; Plectin; 21.
 DR Pfam; PF03501; S10_Plectin; 1.
 DR SMART; SM00033; CH; 2.
 DR SMART; SM00250; PLEC; 33.
 DR SMART; SM00150; SPEC; 4.
 DR PROSITE; PS00019; ACTININ_1; FALSE_NEG.
 DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
 DR PROSITE; PS50021; CH; 2.
 DR Coiled coil: Repeat: Structural protein: Cytoskeleton: Actin-binding;
 KW Phosphorylation; Alternative splicing.
 FT DOMAIN 1 1473 GLOBULAR 1.
 FT DOMAIN 1474 2758 CENTRAL FIBROUS ROD DOMAIN.
 FT DOMAIN 2759 4687 GLOBULAR 2.
 FT DOMAIN 181 406 ACTIN-BINDING.
 FT DOMAIN 185 288 CH 1.
 FT DOMAIN 301 403 CH 2.
 FT REPEAT 646 722 SPECTRIN 1.
 FT REPEAT 743 827 SPECTRIN 2.
 FT REPEAT 840 933 SPECTRIN 3.
 FT REPEAT 1318 1418 SPECTRIN 4.
 FT DOMAIN 1472 1692 COILED COIL (POTENTIAL).
 FT DOMAIN 1724 2760 COILED COIL (POTENTIAL).
 FT REPEAT 2791 2828 PLECTIN 1.
 FT REPEAT 2829 2866 PLECTIN 2.
 FT REPEAT 2867 2904 PLECTIN 3.
 FT REPEAT 2905 2942 PLECTIN 4.
 FT REPEAT 2943 2980 PLECTIN 5.
 FT REPEAT 2984 3018 PLECTIN 6.
 FT REPEAT 3119 3156 PLECTIN 7.
 FT REPEAT 3157 3194 PLECTIN 8.

FT REPEAT 3195 3232 PLECTIN 9.
 FT REPEAT 3233 3270 PLECTIN 10.
 FT REPEAT 3271 3308 PLECTIN 11.
 FT REPEAT 3311 3346 PLECTIN 12.
 FT REPEAT 3348 3385 PLECTIN 13.
 FT REPEAT 3386 3423 PLECTIN 14.
 FT REPEAT 3424 3461 PLECTIN 15.
 FT REPEAT 3462 3500 PLECTIN 16.
 FT REPEAT 3501 3538 PLECTIN 17.
 FT REPEAT 3539 3576 PLECTIN 18.
 FT REPEAT 3577 3614 PLECTIN 19.
 FT REPEAT 3615 3652 PLECTIN 20.
 FT REPEAT 3653 3690 PLECTIN 21.
 FT REPEAT 3691 3728 PLECTIN 22.
 FT REPEAT 3729 3766 PLECTIN 23.
 FT REPEAT 3767 3804 PLECTIN 24.
 FT REPEAT 3805 3842 PLECTIN 25.
 FT REPEAT 3843 3880 PLECTIN 26.
 FT REPEAT 3881 3918 PLECTIN 27.
 FT REPEAT 3919 3956 PLECTIN 28.
 FT REPEAT 3957 3994 PLECTIN 29.
 FT REPEAT 3995 4032 PLECTIN 30.
 FT REPEAT 4033 4070 PLECTIN 31.
 FT REPEAT 4071 4108 PLECTIN 32.
 FT REPEAT 4109 4146 PLECTIN 33.
 FT REPEAT 4147 4184 BINDING TO INTERMEDIATE FILAMENTS.
 FT REPEAT 4185 4222 4 X 4 AA TANDEM REPEATS OF G-S-R-X.
 FT REPEAT 4223 4260 PHOSPHORYLATION (BY CDC2) (BY
 FT REPEAT 4261 4298 SIMILARITY).
 FT REPEAT 4299 4336 MVAGMLPDLQALAIYEVLEFREGVAVAKDRPRSLLPHVP
 FT REPEAT 4337 4374 GYVNLQVMRAMTSLKARGIVRETFAMCHFYWLTNEGIDHL
 FT REPEAT 4375 4412 ROYLHLPELIVPASLQVRVRVAVMVPARRSRPHVOTGCP
 FT REPEAT 4413 4450 LCGPVRGRLPADPDAAREEROYRRKEREAGAPETPVSAT
 FT REPEAT 4451 4488 IYVTLARPGDETPPAT -> MSQQRILRVPEPGLSKRTSS
 FT REPEAT 4489 4526 EDNLVLAIVLASGGR (IN ISOFORM 2).
 FT REPEAT 4527 4564 MVAGMLPDLQALAIYEVLEFREGVAVAKDRPRSLLPHVP
 FT REPEAT 4565 4602 GYVNLQVMRAMTSLKARGIVRETFAMCHFYWLTNEGIDHL
 FT REPEAT 4603 4640 ROYLHLPELIVPASLQVRVRVAVMVPARRSRPHVOTGCP
 FT REPEAT 4641 4678 LCGPVRGRLPADPDAAREEROYRRKEREAGAPETPVSAT
 FT REPEAT 4679 4716 IYVTLARPGDETPPAT -> MPPSGSLPSLTVYGVHVSIA
 FT REPEAT 4717 4754 AVWHMRKCHHQAODEQ (IN ISOFORM 3).
 FT REPEAT 4755 4792 MVAGMLPDLQALAIYEVLEFREGVAVAKDRPRSLLPHVP
 FT REPEAT 4793 4830 GYVNLQVMRAMTSLKARGIVRETFAMCHFYWLTNEGIDHL
 FT REPEAT 4831 4868 ROYLHLPELIVPASLQVRVRVAVMVPARRSRPHVOTGCP
 FT REPEAT 4869 4906 LCGPVRGRLPADPDAAREEROYRRKEREAGAPETPVSAT
 FT REPEAT 4907 4944 IYVTLARPGDETPPAT -> DVNSGSSGSPSPDPLPMNG
 FT REPEAT 4945 4982 KTORSRSGGSGVNGSVLDPAERAVITRIA (IN
 FT REPEAT 4983 5020 ISOFORM 4).
 SQ SEQUENCE 4687 AA; 533527 MW; 9966CAFF71B929751 CRC64;
 Query Match 6.7%; Score 127.5; DB 1; Length 4687;
 Best Local Similarity 21.4%; Pred. No. 13; Mismatches 78; Conservative 51; Indels 109; Gaps 13;
 Matches 78; Conservative 51; Mismatches 127; Indels 109; Gaps 13;
 25 QEELNRPFGAEERVPEDSRWOSRAFPOLGGRPGEGSLSESOPPLIQACPESSC 84
 1755 QQQAELER---AREEAERLEERWQLKANELRLR-----LQAEVAQQS 1796
 85 LRGEKGQNGDSSAGGPPPAEVEPTPAELLAOPCH-----DSEAS 128
 1797 LAQADAEKQEEAEAREARRRKAEDQAVRQELAEQELKOROLTEGTAQOIRLAEOEL 1856
 129 KLGAPAAAGGEPEMGQOOO-----LQKKHRRRPSKKKRWKYYVLTMEKKKPFEEKS 183
 1857 RLRA-----ETEGEHQROLLEELARLDQENATAATQKRO-----ELEELAKVRAEMEV 1906
 184 LRASRTAEMFAKQGVADYNTTQFLMDHDQEPDLTKGLYSKRA-----AA 231
 1907 LLSKARAEESRS-----TSEKSKQRLAEAGRGRELAEARLRALAEAR 1954
 232 KSDDTDDDDPMEGEGEDGSDGMDGDSFLLQDRDSEFYERYHTESLQNMKSQELIKCY 291
 1955 RHRELAEEADAARQRAEADG-----VLTEKLAISQATRLK-- 1989

```

OY 292 LELEKLSRMEDENNRLLESKRLGGDA-RVRELELELDRLAE-----NQLLTENE 344
DB 1990 TEAELALKEAEKERERL-----RLAEDEAFORRLEREOAHOHRADEIRLAQLRKASESE 2045
OY 345 LHRQO 349
DB 2046 LEROK 2050

RESULT 14
TRHY_RABIT
ID TRHY_RABIT STANDARD: PRT: 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Trichohyalin.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: Z19092; CAA79519.1;
CC PIR: S28589; S28589.
CC HSSP: P02633; 41CB.
CC InterPro: IPR001751; CABP_S100.
CC InterPro: IPR002048; EF-hand.
CC Pfam: PF00036; efhand.1.
CC Pfam: PF01023; S_100.1.
CC ProDom: PD003407; CABP_S100.1.

```

```

DR PROSITE: PS00018; EF_HAND.1.
DR PROSITE: PS00303; S100_CABP.1.
KW Keratinization; Repeat; Calcium-binding.
FT DOMAIN 1 91 S-100 LIKE.
FT CA_BIND 22 33 EF_HAND 1 (LOW AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 EF_HAND 2 (HIGH AFFINITY) (POTENTIAL).
SQ SEQUENCE 1407 AA: 183781 MW: 461702.159f12B/F CRC64;

Query Match
Best local Similarity 23.7%; Score 126.5; DB 1; Length 1407;
Matches 83; Conservative 48; Mismatches 124; Indels 95; Gaps 17;

OY 25 QEE-----LMPERPGEAEERVPEDSRWOSRAFPLQGRPGEGSLESOPPLQTOAC 79
DB 349 OEERRROLLAEVEEQAREGESLTPRMQLOESFAGAR-----QSKVSRPRGEGS- 402
OY 80 PESSCLREGKCGNGDSSAGSDPFPFAVEVLPPEVLAQPCHDSEAKLGAPAGGEE 139
DB 403 -----LRQDERROER-----ERELEBO-----ARRQO 427
OY 140 EW-GGGOOROLGKKHRRPSKKRHMKPYKLTWEKKRPDEKOSLRASRAEMPAKQ 198
DB 428 QMGAESEERRRRLSARPSLREROLRAER--QEOQPRFEEDQRRR-RQL----- 479
OY 199 PVAPYNTTQFLMDHDQEEPLDKTGLYSKRAAKSDTSDDEMEGEGEGSDGMDGD 258
DB 480 -----QFL-----EEEBOL-----QRERRAQOLOEEDSPQEDRRRRRQEQORPCQ 520
OY 259 GSEFLQDFSEYERYHT-----ESLQ-NMSKOLLEYLEKLSRMD 303
DB 521 TWRWQLOE--EAORRRRTLYAKPGQEOLEEBELDREKRREERYREEE-LQREED 577
OY 304 ENNRRLLESKRLGGDAVARVELELDRLAENMLQLLTENELH--RQER 351
DB 578 EKRR-RQEREROYRLELELQEEQLNDRKLRREEQLQEEERELRQER 626

RESULT 15
TRHY_SHEEP
ID TRHY_SHEEP STANDARD: PRT: 1549 AA.
AC P22793;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1994 (Rel. 30, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Trichohyalin.
GN THH.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-93260018; PubMed-7684041;
RX Fietz M.J., McLaughlin C.J., Campbell M.T., Rogers G.E.;
RT Analysis of the sheep trichohyalin gene: potential structural and
RT calcium-binding roles of trichohyalin in the hair follicle.";
RL J. Cell Biol. 121:855-865(1993).
RN [2]
RP SEQUENCE OF 1016-1549 FROM N.A.
RC STRAIN-Merino-Dorset horn x Border Leicester; TISSUE-Wool follicles;
RX MEDLINE-90130632; PubMed-2298812;
RA Fietz M.J., Presland R.B., Rogers G.E.;
RT "The CDNA-deduced amino acid sequence for trichohyalin, a
RT differentiation marker in the hair follicle, contains a 23 amino acid
RT repeat.";
RL J. Cell Biol. 110:427-436(1990).
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:21:00 ; Search time 26 Seconds
(without alignments)
406.263 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPPLSEYQHQPOTSNGT.....LTENELHROGERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents.AA.*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/PCYUS.COMB.pep.*
7: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	7.0	538	4	US-09-309-572-12 Sequence 12, Appl
2	133	7.0	1737	4	US-09-309-572-13 Sequence 13, Appl
3	131	6.9	1898	1	US-08-056-200-94 Sequence 94, Appl
4	131	6.9	1898	2	US-08-800-644-94 Sequence 94, Appl
5	125	6.5	754	4	US-09-214-564A-2 Sequence 2, Appl
6	123.5	6.5	557	4	US-08-979-608A-5 Sequence 5, Appl
7	123	6.4	1829	4	US-09-157-420-1 Sequence 1, Appl
8	122	6.4	723	1	US-07-814-964-11 Sequence 11, Appl
9	122	6.4	723	1	US-08-258-442-11 Sequence 11, Appl
10	122	6.4	723	1	US-08-328-809-6 Sequence 6, Appl
11	122	6.4	723	1	US-08-328-809-6 Sequence 6, Appl
12	121	6.3	538	4	US-09-370-368-9 Sequence 9, Appl
13	120	6.3	1780	1	US-08-769-309A-5 Sequence 5, Appl
14	120	6.3	1780	3	US-08-994-570-5 Sequence 5, Appl
15	118	6.2	530	4	US-09-562-737-39 Sequence 39, Appl
16	116	6.1	530	4	US-08-979-608A-8 Sequence 8, Appl
17	116	6.1	2972	4	US-09-579-181-1 Sequence 1, Appl
18	116	6.1	3118	4	US-09-579-181-1 Sequence 1, Appl
19	114.5	6.0	1958	1	US-07-945-283-2 Sequence 2, Appl
20	114	6.0	683	6	5210183-3 Patent No. 5210183
21	112.5	5.9	599	3	US-08-556-419-22 Sequence 22, Appl
22	112.5	5.9	629	3	US-08-556-419-23 Sequence 23, Appl
23	112.5	5.9	764	4	US-09-370-838-67 Sequence 67, Appl
24	112	5.8	802	4	US-09-156-316-1 Sequence 1, Appl
25	111	5.8	1162	2	US-08-728-323A-2 Sequence 2, Appl
26	111	5.8	1162	4	US-09-298-568-2 Sequence 2, Appl
27	110	5.8	576	2	US-08-533-306A-2 Sequence 2, Appl

28	110	5.8	576	2	US-08-742-923A-2 Sequence 2, Appl
29	108.5	5.7	657	3	US-08-893-852A-3 Sequence 3, Appl
30	108.5	5.7	657	4	US-08-821-818-3 Sequence 3, Appl
31	108	5.7	671	4	US-09-605-785-380 Sequence 380, App
32	108	5.7	671	4	US-09-439-313-380 Sequence 380, App
33	108	5.7	671	4	US-09-352-616A-380 Sequence 380, App
34	108	5.7	1115	2	US-08-568-459A-2 Sequence 2, Appl
35	108	5.7	1115	2	US-08-487-828B-2 Sequence 2, Appl
36	108	5.7	1115	4	US-09-210-288-2 Patent No. 5198347
37	108	5.7	1115	6	US-09-605-785-378 Sequence 378, App
38	108	5.7	1719	4	US-09-439-313-378 Sequence 378, App
39	108	5.7	1719	4	US-09-352-616A-378 Sequence 378, App
40	108	5.7	1719	4	US-08-149-097D-35 Sequence 35, Appl
41	108	5.7	2509	2	US-09-457-708-2 Sequence 2, Appl
42	107.5	5.6	1164	4	US-09-718-692-2 Sequence 2, Appl
43	107.5	5.6	1637	4	US-09-718-852-2 Sequence 2, Appl
44	107.5	5.6	1637	4	US-09-718-852-2 Sequence 2, Appl
45	107.5	5.6	1637	4	US-09-718-815-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-309-572-12
Sequence 12, Application US/09309572
Patent No. 6440730
GENERAL INFORMATION:
APPLICANT: Heinrich-Pette-Institut
TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
FILE REFERENCE: P50489
CURRENT APPLICATION NUMBER: US/09/309,572
CURRENT FILING DATE: 1998-05-11
EARLIER APPLICATION NUMBER: DE 198 56 463
EARLIER FILING DATE: 1998-11-26
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 538
TYPE: PRT
ORGANISM: Moloney murine leukemia virus
FEATURE:
OTHER INFORMATION: gag protein
US-09-309-572-12

Query Match	7.0%	Score 133:	DB 4:	Length 538:
Best Local Similarity	20.9%	Pred. No. 0.0019:		
Matches	89:	Conservative	40:	Mismatches 142: Indels 154: Gaps 15:
QY	30	PERPPGA-----EERVPEDSRMOSRAFPOLGGRGPE---	CEGS-----LESQPPPL	74
DB	107	PPLPSPAPSLPLEPRKSTPPRSSLYPALT-PSLGAKKPPVLSDSGPIIDLITEDPPV		165
QY	75	QTACPESSCLREGCEGQNDSSAGGDPPEPAEVEFTPEALL---	AQCHDSASKL	130
DB	166	RDRPPPSD---RDG---NGGEATPAGEAP-----DPSPASLRARRREPPVADSTTSGA		214
QY	131	GAPAGGEEEMGOQOOLGKKHRRRPSKKRRWKRYLLTWEEKKPPKOSLRASRLR		190
DB	215	FPLRAGN-----COLQYWPSSSDLYNWK-----NNNPSFSDPKLTALTE		257
QY	191	AEMFAGQVPAYNTQFLMDDHDQEEPLKTQ-----LYSKRAAKGSD-----		235
DB	258	SVLL-----THQPTWDDQOQLGLTLTGEEKORVLLERKAKVRDDGRPTOLPNE		307
QY	236	-----TSDDFMEEGGEF-----	DGSDMGDGGE	261
DB	308	VAAAPLEPRDDYTTQAGRNHLVHYRQLLAGLONAGRSPTNLAKVKGITGPPNPSA		367
QY	262	FLORDFSFYERY-----HTESLONNSKDELKE		290
DB	368	FLER-LKEAYRRTTPDPEDPGQETNVSMSTWOSAPDIGRKLRLDLKNTIGDLVR-		425

```

OY      291 YTELECSRMDENNRLRLEESKRUGODARY-RETELEIDRLAENJOLITNELHROO 349
Db      426 --BAEKIFMKRTEPPER-----EERIRTEKEERRRTEDEOKKEKENDRRHR 472

OY      350 ERAPL 354
Db      473 EMSKL 477

RESULT 2
US-09-309-572-13
: Sequence 13. Application US/09309572
: Patent No. 6440730
: GENERAL INFORMATION:
: APPLICANT: Heinrich-Pette-Institut
: TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
: FILE REFERENCE: P50489
: CURRENT APPLICATION NUMBER: US/09/309,572
: CURRENT FILING DATE: 1999-05-11
: EARLIER APPLICATION NUMBER: DE 198 56 463
: EARLIER FILING DATE: 1998-11-26
: NUMBER OF SEQ. ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 13
:
: LENGTH: 1737
: TYPE: PRT
: ORGANISM: Moloney murine leukemia virus
: FEATURE:
: OTHER INFORMATION: gag-pol protein
: US-09-309-572-13

```

Query Match	7.0%	Score	133;	DB	4;	Length	1737;
Best Local Similarity	20.9%	Pred	No. 0.0088;				
Matches	89;	Conservative	40;	Mismatches	142;	Indels	154;
						Gaps	19;

```

QY PERPCA-----EERPEEDSRMOSRAFPOLGGPGE---GBGS-----LESOPL 74
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db PPLPSAPSLPLEPPNSTPPRSSLYPAL-PSLGAKPKQVULSDGGLIDLTDEPPY 167
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY QTOCAPESCLREGEKONGDDSSAGDPPPAEVEPTAEALL---AOPCHSEASKL 130
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db RDRPPSPD--RDQ-----NCGEATPPAGEAP-----DSPMASRLRGREPVPADSTTQA 214
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 131 GAPAAGEEMEGOOORLOKKKHRRRPPSKKKHMKRYTLTWBEKKKPEKESKLASIR 190
      ||| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 215 PPLRAGN-----GOLOYWPSSSDLYMWK-----NNNPSEDPGKLTALIE 25
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 191 AEMEKAQPVAFYNTTOFLMDHDHOEPDLKTG-----LYSRAAKSD----- 231
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 258 SVLL-----THQPTWDDCQLGLTLLTGEEKRVLEARKAVRDDGRPTQLPNE 30
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 236 -----TSDDPFMEGCEE-----DGSGMGMGDDSE 26
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 308 VDAAPFLERPMDYTTOAGRNLHVHYROLLLAGLONAGRSPTNLAKEYITGPNESPJA 36
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 262 FLOADFSETERY-----HTESLONNSKOELIKE 291
      ||| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 368 FLER-LKEAYRRTYPYPEDPGOETINVSMTFOWSAPDLGRKLEREDLKNTTDIVR- 42
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 291 YLELEKLSRMEDENNRLRESKRFGDDAR-RELELEDLRRAENTLOLTENELHQO 34
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 426 --EAEKIYNKRETPER-----EERIIRTEBEKEERRRTDEDEKEKERDRRRHR 47
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 350 ERAPL 354
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 473 EMSKL 477

```

```

RESULT 3
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:

```

APPLICANT: Steinert, Peter M.
 APPLICANT: Lee, Seung-Chul
 APPLICANT: Kim, In-Gyu
 APPLICANT: Chung, Soo-Il
 APPLICANT: Park, Sang-Chul
 TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
 TITLE OF INVENTION: Methods of Using Same
 NUMBER OF SEQUENCES: 117
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe, Martens, Olson & Bear
 STREET: 620 Newport Center Drive, Sixteenth Floor
 CITY: Newport Beach
 STATE: CA
 COUNTRY: U.S.A.
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/056,200
 FILING DATE: 30-APR-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Fedrick, Michael F.
 REGISTRATION NUMBER: 36,799
 REFERENCE/DOCKET NUMBER: NIH054.001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (714) 760-0404
 TELEFAX: (714) 760-9502
 INFORMATION FOR SEQ ID NO: 94:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1898 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-056-200-94

Query Match	6.9%	Score 131	DB 1	Length 1898
Best Local Similarity	22.4%	Pred. No. 0.015		
Matches	79	Conservative	54	Mismatches 146
				Indels 74
				Gaps 13
QY	31	ERPFAEER-----VPEDSRMOSRAFLPGRGPEGESESLSPAPLQTOACCESS	83	
Db	813	ERQLRAEERQROEOFLPREEEKR-----GRGRERKEKLQFLEEBQLORRBAQ	865	
QY	84	CLREGKCGNGDDSSAGDPFPPPAEVEPTPAE-----LLAOPCHDSESKIGAPAAQ	136	
Db	866	QLOEEEDLDQEOERROEORRDQWMLQLEERKRRNRHTYAKPALQEOLRKQOULLQ	925	
QY	137	GEEMWGQOORQLGKKHRRSRKKKRNHKPPYKLTWEEKKKFDEKOSLRASIRAEWPK	196	
Db	936	EEELQREERE-----KRRQEOEROYREEBQLOOEEQLLREERKRRQOEREROYRK	979	
QY	197	GQVAVYNTQFLMDHDQ---EEFDLTGLYSKRAAKSDTSDDPDMEGEGEDGSD	253	
Db	960	DKK-----LQKEEQLLGEPEKR-----RQERKKYREEBLQOE-----	1016	
QY	254	GMGGDSEFLQDFSETYER-----YHTESLQNNKSOELIKYILEKCLSRMED--EN	305	
Db	1017	-----EEQLLREERKRRQOEWEROYKKKDELQOEEBQLLREERKRRLOEROEROYREE	1070	
QY	306	NRLRLSKRLGDDARV---FLIEL---DRLRKNOLLTEN-ELAROER	351	
Db	1071	ELQOEEBQLGEEERTRRQELEROYKKEEQLQOEEBQLLREERKRROR	1123	

RESULT 4
US-08-800-644-94
Sequence 94, Application US/08800644
Patent No. 5958752
GENERAL INFORMATION:

```

: APPLICANT: Steinert, Peter M.
: APPLICANT: Lee, Seung-Chul
: APPLICANT: Kim, In-Gyu
: APPLICANT: Chung, Soo-Il
: APPLICANT: Park, Sang-Chul
: TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
: TITLE OF INVENTION: Methods of Using Same
: NUMBER OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/800,644
: FILING DATE: 14-FEB-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/056,200
: FILING DATE: 30-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Fedtich, Michael F.
: REGISTRATION NUMBER: 36,799
: REFERENCE/DOCKET NUMBER: NIH054.001A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (714) 760-0404
: TELEFAX: (714) 760-9502
: INFORMATION FOR SEQ ID NO: 94:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1898 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-800-644-94

```

```

Query Match      6.9%; Score 131; DB 2; Length 1898;
Best Local Similarity 22.4%; Pred. No. 0.015;
Matches 79; Conservative 54; Mismatches 146; Indels 74; Gaps 13;

```

```

QY 31 ERPPGAEER-----VPEEDSRMOSRAFPOLGRRPCEGSGLESQPPPLQTOACPBSS 83
   |||||
DB 813 EROLRAEEROERQRFLEPEEEKOR-----GRORREKELQFLEEEOLQRRERAQ 865
   |||||
QY 84 CLREGEKGONDDSSAGDFPPPAVEPTPEAF-----LLAOPCHDSEASKLGADPAAG 865
   |||||
DB 866 QLOEEEGLOEDORRRROEOBRODKWMOLEEEFKRRHNTLYAKPALQOLRKQOLLOE 925
   |||||
QY 137 GEEFWGQOQOLGKKHRRRPSKKRHKPYKLTWEEKKFKDEKOSTIRASIRAEAFK 196
   |||||
DB 926 EEEELQREERE-----KRROEOEROYREEEQLOEEEOQLREREKRRROEROYRK 979
   |||||
QY 197 GQAPAPYNTQFLMDHDQ---FEEDLKTGLYSKRAAKSDTSDDDFMEGGEGDGS 253
   |||||
DB 980 DKR-----LQCKEQLLGEPEERK-----RROEREKKYREEEELQEE----- 1016
   |||||
QY 254 GMGDGSSEFLQDSEFYER-----YHTESLQNMKSQELIKYLELEKCLSRMED--EN 305
   |||||
DB 1017 -----EEOILLRERERKRRQEMEROYRKDELQOEEEOQLREREKRRLOEREROYREE 1070
   |||||
QY 306 NRLEFKRLGGDDARV---ELELEL---DLRAENLQLTEN-ELARQOER 351
   |||||
DB 1071 EELQOEEEOQLGERETRRROELEROYRKDELQOEEEOQLREREPEERKRRQER 1123

```

RESULT 5
US-09-214-564A-2

```

: Sequence 2, Application US/09214564A
: Patent No. 6150515
: GENERAL INFORMATION:
: APPLICANT: Sharp, Phillip A.
: APPLICANT: Zhou, Qiang
: TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional
: TITLE OF INVENTION: Elongation By HIV-1 TAT
: FILE REFERENCE: M0656/7042
: CURRENT APPLICATION NUMBER: US/09/214,564A
: PRIOR FILING DATE: 1999-08-18
: PRIOR APPLICATION NUMBER: US 60/021,218
: PRIOR FILING DATE: 1996-07-03
: PRIOR APPLICATION NUMBER: US 60/033,152
: PRIOR FILING DATE: 1996-12-13
: PRIOR APPLICATION NUMBER: PCT/US97/11713
: PRIOR FILING DATE: 1997-07-03
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 754
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-214-564A-2

```

```

Query Match      6.5%; Score 125; DB 4; Length 754;
Best Local Similarity 21.5%; Pred. No. 0.014;
Matches 82; Conservative 50; Mismatches 166; Indels 84; Gaps 17;

```

```

QY 9 YOHOPOTSNCGAAYOEELNPPRPGAEEVPEEDSRMOSRAFPOLG--RCPREGSG 66
   |||||
DB 401 FSEHPSTSKMNA-----QETATGMARFEEPIDEKKFKF---TEDGGEFEESASENNA 448
   |||||
QY 67 LESQPPPLQOQAC-----ESSCLREGEKKGONDDSSAGGPPPAVEPTPEAF---LL 118
   |||||
DB 449 KESSPEKEAEBCGCKESBEGCRKRGEGSCQKESSEGN-PRGSEEDSPKESKKTLL 507
   |||||
QY 119 AOPCHDSEASKLGADPAAGGEEWGOOQOLGKKHRRRPSKKRHKPYKLTWEEKKF 178
   |||||
DB 508 KNDCEENGKLAKESDDLINKSE-----EVGPTKESSEDSSEK-----ESQEDC 551
   |||||
QY 179 DEKOSLRASIRAEAFKQAPVAPYNTQFLMDHDQ--EPDLKTGLYSKRAAKSDT 236
   |||||
DB 552 SEKQSEDSEREFEF-----ENGLKXLDDEGSEKELHENVLKE--LEENDS 596
   |||||
QY 237 SDDFMEEGGEFGDSQMGDGSSEFLQDSEFYER-----TYER-YHTESLQNMKSQ 285
   |||||
DB 597 ENSEF-----EDGSEKVLDEEGSE---REFDESDKEEEDDYERKVFDESDSEKDEE 648
   |||||
QY 286 ELIKYLELEKCLSRMEDENNRLRLLESKRRLGDDARVRELELELRLRAE---NLQLTTE 342
   |||||
DB 649 YADKRGLEADKKRAEBBDADKLFEEESDQKDEDELADGKVEVDADKLFEDDDSNKLFDE 708
   |||||
QY 343 NE-----LHROERAPLSKFG 358
   |||||
DB 709 EEDSSEKLFDDSDERGTGFGFG 730

```

```

RESULT 6
US-08-979-608A-5
: Sequence 5, Application US/08979608A
: Patent No. 6355451
: GENERAL INFORMATION:
: APPLICANT: Lees, Ann M.
: APPLICANT: Lees, Robert S.
: APPLICANT: Law, Simon W.
: APPLICANT: Arjona, Anibal A.
: TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
: BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
: TREATING ATHEROSCLEROSIS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street

```

CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 FILING DATE: 26-NOV-1997
 APPLICATION NUMBER: US/08/979,608A
 FILING DATE: 26-NOV-1997
 APPLICATION NUMBER: US 60/048,547
 FILING DATE: 03-JUN-1997
 APPLICATION NUMBER: US 60/031,930
 FILING DATE: 27-NOV-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Myers, Louis
 REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 557 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-08-979-608A-5

Query Match 6.5%: Score 123.5; DB 4; Length 557;
 Best Local Similarity 22.6%: Pred. No. 0.013;
 Matches 91; Conservative 53; Mismatches 123; Indels 135; Gaps 20;

20 GAAVQELNPERPPGAEEVPE-----EDSRWQRAFPOLGRRPPEGEGSLSSQ 70
 9 GAA--KQPNKSSPGQPEAGAGQGRPPAPAREAGASSQAPGR--PEGAAQAKYAQ 63
 71 PPPL-----QTAQPESSCLREGEKQNGDDSSAGDPFPAEVEPTPEALLAQP 121
 64 PGLADVSEELSRQLEDILTYCV--DNNQAGPGEDEVG-----EP--PEP----- 107
 122 CHDSASKL-----GAPAAAGEEEMGQQ-----POLGKKRRRRRPSKKRR 162
 108 --DAKSRAYVARNGEPEPTPVVNGEKEKTSKAERGTEIITSDEVGDRDHRRRQEKKA 165
 163 HW-----KPYKLTWEKKKKFDEKQSLRASIRAFKAGQAPVAPYNTT 206
 166 KGLCKEITLMTNTLSTPEEKLAALCKKYAEILLEHRNSQKMKLLQKKQ-----S 218
 207 QELMDHDEPPDKTGLYSKRAAKSDDTS-----DDDFMEEGEEDGGSDMGMD 258
 219 QLV-----QEKDILRGHSAIILARSLSELCRELORHNSLKEEG----- 259
 259 GSEFLQRFSELYERHTESLQNSKQELIKLEYLEKLSRMEDENNRLFLSKRLGCD 318
 260 ---VQRRAREEERKKETVSHFQMTLNDI---QLQME---QHNERNSKLQENMELA-- 306
 319 DARVRELELDRLRAENL-----QLTENEHLRQOE 350
 307 -ERLKL-IEQYELREHIDKVFKKHKLQOQLVDAKIQAOE 346

RESULT 7
 US-09-157-420-1
 Sequence 1, Application US/09157420
 Patent No. 6180760
 GENERAL INFORMATION:
 APPLICANT: TAKAI, Yoshimi

APPLICANT: NAKANISHI, Hiroyuki
 APPLICANT: MANDAI, Kenji
 APPLICANT: WADA, Manabu
 APPLICANT: OBAISHI, Hiroshi
 TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-ARADIN"
 FILE REFERENCE: 98-1042*/LC(WMC)/653
 CURRENT APPLICATION NUMBER: US/09/157,420
 CURRENT FILING DATE: 1998-09-21
 NUMBER OF SEQ ID NOS: 1
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO: 1
 LENGTH: 1829
 TYPE: PRT
 ORGANISM: rat
 US-09-157-420-1

Query Match 6.4%: Score 123; DB 4; Length 1829;
 Best Local Similarity 19.9%: Pred. No. 0.068;
 Matches 83; Conservative 59; Mismatches 157; Indels 118; Gaps 15;

9 YOHQFQTSNCGAAV-----QELNPERPPGAEEVPE-----EDSRWQRAFP 53
 1259 YEEKPHMTESDHSIAIQRVTRSOELREKVVQLERHVESGMDRKCDSQMTNOS-- 1316
 54 QLGRRPPEGEGSLSSQPPPIQTAQPESSCLREGEKQNGDDSSAGDPFPAEVEPTP 113
 1317 ---SSVESSSTSSQEHNLHNSKSVTPASTILTKSP-----GRMKTAAVILPTP 1360
 114 EAELLAPCH-----DSEASKIGADPAAGEEEMGQQOROLGKKHRR 155
 1361 VA--VSQPIRTDLPPPPPPAHYTSDFGISMDELPLPPPPANA-AQSAQVAAAEKK 1417
 156 PPSKKRRHWPYKLTWE-EKKKPEKQSLRASIRAFKAGQVPA----- 201
 1418 REEHQRWYEEKARLEEREKRRQEKQLQMTQSLNPASFPLAQAPKRPSTLQR 1477
 202 PYNTQFLMDHDEEP-----DLKTGLYSKRAAKSDDTSDDDFMEEGEEDGGSDGM 255
 1478 PQETV--IRELQPOQPPRTERRDLQYTTISKEELSDGSLSPBW----- 1521
 256 GGDSEFLQRFSELYERHTESLQNSKQELIKLEYLEKLSRMEDENNRLR----- 309
 1522 ---KRAREKLEKQQGMHIVDM---LSKEIHQLQKKGRTAESDRLKRLMEW 1569
 310 ---LESKRLGDDDAKAVRELE---LDRLAENLQLTENEHLRQOERAPLSP 356
 1570 QFQKRLQSSKKKDDDEEDDDVDYTMILMGRLEARRARLQDEERRQQLLEKMK 1626

RESULT 8
 US-07-814-964-11
 Sequence 11, Application US/07814964
 Patent No. 5359047
 GENERAL INFORMATION:
 APPLICANT: Donahue, Brian A.
 APPLICANT: Toney, Jeffrey H.
 APPLICANT: Bruhn, Suzanne L.
 APPLICANT: Pili, Pieter M.
 APPLICANT: Brown, Steven
 APPLICANT: Kellett, Patti
 APPLICANT: Essigmann, John M.
 APPLICANT: Lippard, Stephen J.
 TITLE OF INVENTION: DNA Structure Specific Recognition
 NUMBER OF INVENTIONS: Protein and Uses Therefor
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: 2 Militia Drive
 CITY: Lexington
 STATE: MA
 COUNTRY: USA
 ZIP: 02173
 COMPUTER READABLE FORM:


```

: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/11107
: FILING DATE: 19921218
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/539,906
: FILING DATE: 18-JUN-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: MIT-4787AAA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 723 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Drosophila melanogaster
: IMMEDIATE SOURCE:
: CLONE: Drosophila SSRP (predicted)
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 458..507
: OTHER INFORMATION: /label= Acidic
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 518..547
: OTHER INFORMATION: /label= Basic I
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 547..620
: OTHER INFORMATION: /label= HMG-box
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 632..649
: OTHER INFORMATION: /label= Basic II
: NAME/KEY: Domain
: LOCATION: 657..723
: OTHER INFORMATION: /label= Mixed Charge
PCT-US92-11107-11
Query Match 6.4%; Score 122; DB 5; Length 723;
Best Local Similarity 20.4%; Pred. No. 0.025;
Matches 61; Conservative 55; Mismatches 125; Indels 58; Gaps 9;
QY 63 GEGLESGPPPIQTQACPESSCLREGEKGNQGDSSAGGPPPP--PAVEVPTPEALLA 119
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 GDSQNEPEPAYIARLAEAREKEDEDGSDSESTDEDFKPNENSDVAEEYDSNVS 500
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 QPCHDSEASKLGAPAGGEEMGQQOOLGKKHRRRPSKKRMKPPYKLTWEKKRPD 179
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 DSDDDSDAS-----GGGSDSGAKKKKE--KKSEKKKKKKKKH-KEKERTKKSKKKD 551
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 EKQSLRSRIRAEAFAGQVPAYNTTQFLMDHD-----QEPDLKTLGYSKR----- 228
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 552 SGKPKRA-----TTAFMLMLNDTRESIKRENPGLKVIETIAKGGEMW 593
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 229 -----AAKSDTSDDDPFMEEGEGEDGSDGMSFLLORDSETTERKYTE 277
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 594 KELKDKSKMEDAAKKDKORYHDEMRNTKPEAGGSDNEKGKSSKKRRKTEPSPSKANTS 653

```

```

QY 278 SLQNMRSKOELIKEYLELEKCLSRMEDENNR-LRLESKRLGGDARVELLELDRLRAE 335
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 654 GSGFKS-----KEYISDDSTSSDDEKDNERRAKKSNPPSDGDAKKKAKSESPEESE 707
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 12
US-09-370-368-9
: Sequence 9, Application US/09370368
: Patent No. 6258932
: GENERAL INFORMATION:
: APPLICANT: Anders Vahine
: TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY
: FILE REFERENCE: TRIPEP.003A
: CURRENT APPLICATION NUMBER: US/09/370,368
: CURRENT FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 9
: LENGTH: 538
: TYPE: PRT
: ORGANISM: Moloney Murine Leukemia Virus
US-09-370-368-9
Query Match 6.3%; Score 121; DB 4; Length 538;
Best Local Similarity 21.3%; Pred. No. 0.02;
Matches 86; Conservative 49; Mismatches 147; Indels 122; Gaps 19;
QY 30 PERPPGAERYPEEDSRWQSRAPPOLGRRPPE--GEGS-----LESQPPPIQTQACPE 81
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 PLPPPLS---TPPOSSLYPALT-PSLGAKPKPYLSDSGPLDLIDLEDPPIYDRPPPP 172
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 SSCIREGEKGNQGDSSAGGDFPPPAVEVPTPEALL-----AOPCHDSEASKLGAPAGG 137
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 SD--RDGDSG---EATPAGEAP-----DSPMASRLGRREPPVASTTSQAFPLRTGG 221
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 138 EEWGQQOOLGKKHRRRPSKKRHKMPPYKLTWEKKKFKDEKOSLRASRIRAEKMPAGK 197
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 N-----GQIQVYFPFSSDLYNKK-----NNNPSSEDPGKLTALIESVLL--- 261
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 198 QVPAVYNTTQFLMDHDQEEPDLTG-----LYSKRAAKSSD----- 235
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 262 -----THOPTWDDCOLLGLTLTGEEKORVLLKARKAVAGDDGRRPTQLPNEYDAFPL 314
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 -TSDDDMEEGGER-----DGSDGMSGSGSEFLORDFS 268
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 ERPMWEYTTQAGRNHLVYRQLITAGLQNAGRSPTNLAKYKGLTQGNESPSAFIER-LK 373
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 ETVRYHYTESLQ-----NMSKOELIKEYLELEKCLSRMEDENNRL-----RLSK 313
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 EAVRYTPYPDEPDGQETNVSMTFINSAPDIGKRLERLDLRKTKIGDLVREAFERLFNK 433
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 RLGGD--DARV-RELELDRLRAENLQLTENELHROQERAPL 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 434 RETPEEREERIRREKEERREKREDEQKEKERDRRRHREMSRL 477
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 13
US-08-769-309A-5
: Sequence 5, Application US/08769309A
: Patent No. 5741890
: GENERAL INFORMATION:
: APPLICANT: Scott, John D.,
: APPLICANT: Nauert, Brian J.,
: APPLICANT: Klauk, Theresa M.,
: TITLE OF INVENTION: Protein Binding Domains of Gravin
: NUMBER OF SEQUENCE: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower/233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America

```



```
Db 307 SESEPEPELHETPRRPAFLPVYODDTNSEYEMWSESEPDLSYDADSPWLLSA--LVSR 363
QY 59 PCPEGE-----GSLESQPPLOTQACPESSCLREGCKGONGDSSAGDF----- 103
Db 364 MISEGDSPTKCPGOCESPAFRL-----PEFFASQANSVPQCCODPEAGPHHELVDMDTLC 418
QY 104 --PPPA-----EVEPTPEALLAQPCHDSEASKLQAPAGGEEMWGOQOQOLGKKKRR 154
Db 419 IPPPPAAPAAKLGPAQGP-CMFLSNPTRTNTPLMATP-----GRQARP-GRCSA 468
QY 155 RRPSSKKRRHKPYKLTWEKKKFKDEKQSLRA-----SRIAEKFAKGQPVAP---- 202
Db 469 RCSEEE-----EEDESEDEDEEDATDVVPPGSRVYGS-----TAPLDW 509
QY 203 --YNTTQ--FLMDHDOEEPDLTGLYSKRAAKSDOTSDDDFMEEGGEEDGSDGMGD 258
Db 510 LVYDAVKYTYVDEHTOLE-----AVSLRRCAGLNDSEEDSSCFASEEAGATGLGSD 563
QY 259 G-----SEFLQRDSEETERYHTESLQNM-----KOELIKEYLELEKCLS 299
Db 564 QVEDHSPDSPDLTFIKKFLNVFVNKTSRSSSTESLGLFSCVXNGMEREQTHRAVNRFTP 623
QY 300 RMEDENNRLRLKESKRUGDDARYELE 326
Db 624 RHPDE-----QELDVDDPVLEAE 642
```

Search completed: June 17, 2003, 10:25:34
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 17, 2003, 10:55:42 : Search time 194 Seconds
(without alignments)
2679.659 Million cell updates/sec

Title: US-09-972-758A-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYOHQPTSNCTG.....LTENELHROQERAPLSKFGD 359

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US09972755/runat_17062003_095239_24438/app.query.fasta_1.519
-DB=publshed_Applications_MA -QFMT=fastap -SUFFIX=rrpb -MIMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -ONITS=bits -START=1 -END=1 -MATRIX=biosum62
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=ext -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09972758.@CGN.1.1.57 -runat 17062003_095239_24438
-NCPU=6 -ICPU=3 -NO_MMAR -LAREQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : Published_Applications_MA:*

1: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2.6/ptodata/1/pubpna/PCT_NEM_PUB.seq:*
3: /cgn2.6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2.6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2.6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2.6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2.6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2.6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14: /cgn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	2199	10	US-09-745-763-33
2	1902	99.6	1080	9	US-09-972-758-1
3	691	36.2	414	10	US-09-983-965-4895
4	629	32.9	461	9	US-09-954-531-813

	C	5	619	32.4	495	9	US-09-918-995-403	Sequence 403, App
		7	184	9.6	549	9	US-09-991-936-1712	Sequence 1712, Ap
		6	152	8.0	2421	10	US-09-974-300-7891	Sequence 7891, Ap
		8	147.5	7.7	3773	10	US-09-925-302-47	Sequence 47, Appl
		9	147	7.7	7453	9	US-10-037-270-248	Sequence 248, App
		10	147	7.7	7501	9	US-10-037-270-249	Sequence 249, App
		11	145	7.6	4301	10	US-09-917-800A-1666	Sequence 1666, Ap
		12	141	7.4	2962	9	US-09-291-417-106	Sequence 106, App
		13	139	7.3	6537	9	US-10-175-523-186	Sequence 186, App
		14	138.5	7.3	5835	9	US-09-927-597-1	Sequence 1, Appl
		15	138.5	7.3	5937	9	US-09-927-597-3	Sequence 3, Appl
		16	138.5	7.3	6861	9	US-10-171-311-161	Sequence 161, App
		17	138.5	7.3	6900	9	US-10-171-311-163	Sequence 163, App
		18	138	7.2	1960	10	US-09-834-975-833	Sequence 833, App
		19	137	7.2	8332	10	US-09-006-298-1	Sequence 1, Appl
	C	20	135.5	7.1	2654	9	US-10-050-704-40	Sequence 40, Appl
		21	135.5	7.1	3476	9	US-10-245-103-51	Sequence 51, Appl
		22	135.5	7.1	3476	9	US-10-245-107-51	Sequence 51, Appl
		23	135.5	7.1	3476	9	US-10-245-143-51	Sequence 51, Appl
		24	135.5	7.1	3476	9	US-10-245-771-51	Sequence 51, Appl
		25	135.5	7.1	3476	9	US-10-245-851-51	Sequence 51, Appl
		26	135.5	7.1	3476	9	US-10-245-883-51	Sequence 51, Appl
		27	135.5	7.1	3476	9	US-10-237-535-51	Sequence 51, Appl
		28	135.5	7.1	3476	9	US-10-238-183-51	Sequence 51, Appl
		29	135.5	7.1	3476	9	US-10-238-283-51	Sequence 51, Appl
		30	135.5	7.1	3476	9	US-10-238-370-51	Sequence 51, Appl
		31	135.5	7.1	3476	9	US-10-245-055-51	Sequence 51, Appl
		32	135.5	7.1	3476	9	US-10-245-147-51	Sequence 51, Appl
		33	135.5	7.1	3476	9	US-10-245-730-51	Sequence 51, Appl
		34	135.5	7.1	3476	9	US-10-245-739-51	Sequence 51, Appl
		35	135.5	7.1	3476	9	US-10-246-210-51	Sequence 51, Appl
		36	135.5	7.1	3476	9	US-10-239-196-51	Sequence 51, Appl
		37	135.5	7.1	3476	9	US-10-243-024-51	Sequence 51, Appl
		38	135.5	7.1	3476	9	US-10-243-409-51	Sequence 51, Appl
		39	135.5	7.1	3476	9	US-10-245-033-51	Sequence 51, Appl
		40	135.5	7.1	3476	9	US-10-245-621-51	Sequence 51, Appl
		41	135.5	7.1	3476	9	US-10-245-880-51	Sequence 51, Appl
		42	135.5	7.1	3476	9	US-10-243-095-51	Sequence 51, Appl
		43	135.5	7.1	3476	9	US-10-245-185-51	Sequence 51, Appl
		44	135.5	7.1	3476	9	US-10-245-427-51	Sequence 51, Appl
		45	135.5	7.1	3476	9	US-10-245-473-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1
US-09-745-763-33
Sequence 33, Application US/09745763
Patent No. US20020065394A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth

McCoy, John M.
Lavallee, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merbery, David
Trecay, Maurice

Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 219
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 2199 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: <Unknown>
US-09-745-763-33

Alignment Scores:
Pred. No.: 1 22e-167 Length: 2199
Score: 1910.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-972-758a-2 (1-359) x US-09-745-763-33 (1-2199)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnGly 20
DB 707 ATGGCCGAGCCATTCTTGTACAGATATCAACACGACCTCAAACTGACACTGACAGGT 766
QY 21 AlaAlaAlaValGlnGlnGlnGlnLeuAsnProGluArgProProGlyAlaGlnGlnVal 40
DB 767 GCGTGGCTGTCCAGGAAGAGCTGAACCTGAGCGCCCGGCGGAGGAGGAGGCGG 826
QY 41 ProGluGluAspSerArgTTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
DB 827 CCGAGAGGAGACGATGAGTGGCATGGAAGCGCTTCCCGCAGTGGGTGGCCCTCGGG 886
QY 61 ProGluGlyGlnGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaGlyPro 80
DB 887 CCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 946
QY 81 GluSerSerCysLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
DB 947 GAATCTAGCTGCTTGAAGAGAGGCGGAGAAAGGCGCAATGAGGAGGAGGAGGAGGAG 1006
QY 101 GlyAspPheProProProAlaGlnValGluProThrProGluAlaGlnLeuLeuAlaGln 120
DB 1007 GCGGACTTCCCGCGCGCGCGCGAGAGTGAACCGAGCGCGCGAGCTGCTCGCGCCAG 1066
QY 121 ProGlyHisAspSerGlnAlaSerLeuGlyAlaProAlaAlaGlnGlyGlnGlnGln 140
DB 1067 CCTTGTCATGACTCCAGGCGCAGTAACTTGGGGGCTCCCGCGCAGGCGGCGCAAGAGAG 1126
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGln 160
DB 1127 TGGGAGACAGACAGACAGACAGCTGGGGAGAAAGAAACATAGGAGAGCGCCGCTCAAGAG 1186
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGlnGlnLysLysLysPheAspGlu 180
DB 1187 AAGCGGCATTTGGAACCCGTAACAAAGCTGACCTGGGAGAGACAAAGAAAGTTCCGACGAG 1246
QY 181 LysGlnSerLeuArgAlaSerArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 200
DB 1247 AAACAGAGGCTTGGAGCTTCAAGAGATCCGAGCCGAGATGTTGCCCAAGGGCGACGCGTTC 1306
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnGlnGlnGln 220
DB 1307 GGGCCCTATACACACGAGCTTCTCATGATGATACGACGACGAGAGAGCGCGGATCTTC 1366
```

```

QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240
DB 1367 AAACCGGCGCTTACTTCCAGAGCGGGCGCCGCCCAATCCGACGACACGACGATGACGAC 1426
QY 241 PheMetGlnGlnGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
DB 1427 TTCATGGAAGAGAGGGGTGAGGAGGATGGGGGACAGGATGGGATGGAGAGGAGCGGACG 1486
QY 261 GluPheLeuGlnArgAspPheSerGlnTyrTrpGluArgTyrHisThrGlnSerLeuGln 280
DB 1487 GAGTTTCTCCACCGGACTTCTCGGAGAGCTGAGACGGGTACACACGAGAGCGCTGCGG 1546
QY 281 AsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
DB 1547 AACATGACGACAGCAGAGCTCATCAAGAGAGTACCTGGAACCTGAGAGAGCGCTTCCGCC 1606
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
DB 1607 ATGAGAGAGCAGACACACCGGCTGGGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAG 1666
QY 321 ArgValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
DB 1667 CGTGGCGGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGAGCTGAGAGCTGAG 1726
QY 341 ThrGluAsnGlnLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
DB 1727 ACCGAGACGAGACTGACCGCGGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1783

RESULT 2
US-09-972-758-1
: Sequence 1, Application US/09972758
: Patent No. US20020160497A1
: GENERAL INFORMATION:
: APPLICANT: Case Western Reserve University
: APPLICANT: Montano, Monica
: APPLICANT: Wiltman, Bryan
: TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth
: FILE REFERENCE: 27708/04004
: CURRENT APPLICATION NUMBER: US/09/972,758
: CURRENT FILING DATE: 2001-10-05
: PRIOR APPLICATION NUMBER: US 60/238,187
: PRIOR FILING DATE: 2000-10-05
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 1
: LENGTH: 1080
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-972-758-1

Alignment Scores:
Pred. No.: 2 99e-167 Length: 1080
Score: 1902.00 Matches: 357
Percent Similarity: 99.72% Conservative: 1
Best Local Similarity: 99.44% Mismatches: 1
Query Match: 99.58% Indels: 0
DB: 9 Gaps: 0

US-09-972-758a-2 (1-359) x US-09-972-758-1 (1-1080)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnGly 20
DB 1 ATGGCGGAGCCATTCTTGTCAAAATATCAACACGACCTCAAACTGACACTGACAGGT 60
QY 21 AlaAlaAlaValGlnGlnGlnGlnLeuAsnProGluArgProProGlyAlaGlnGlnVal 40
DB 61 GCTGCTGCTGTCCAGGAAGAGCTGAACCTGAGCGCCCGGCGGAGGAGGAGGAGGAGGAG 120
QY 41 ProGluGluAspSerArgTTrpGlnSerArgAlaPheProGlnLeuGlnGlyArgProGly 60
DB 121 CCGGAGAGGAGACGATGAGTGGCATGGAAGCGCTTCCCGCAGTGGGTGGCCCTCGGGG 180
```

QY 61 ProGluGluGluGluGluSerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
Db 181 CCGGAGGGGGAGAGGAGCGCTGGAAATCCCAACACCTCCCTTGAGACCCAGCGCTGTCCA 240
QY 81 GluSerSerCysLeuArgGluGluGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
Db 241 GAATTTACTCTGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
QY 101 GlysPhePheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
Db 301 GGCGACTTCCCGCGCGCGCGAGAAAGTGGAAACCGACGCCGCGAGCGCTCTCTCCCGCAG 360
QY 121 ProCysHisAspSerSerGluAlaSerLysLeuGlnLysAlaProAlaAlaGluGluGlu 140
Db 361 CTTGTCTAGTACTCCGAGCGCCAGTAACTTGGGGGCTCTCTCCGCGAGGGCGGAGAGAGAG 420
QY 141 TrpGluGlnGlnGlnArgGlnLeuGlnLysLysLysHisArgArgArgProSerLysLys 160
Db 421 TGGGAGACGCGAGAGACAGACCTGGGAGAGAAAACATAGAGAGACCGCTCCAGAGAG 480
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysPheAspGlu 180
Db 481 AAGCGGATTTGGAAACCGTACTACACCTGGAAGTGGAGAGAGAGAGAGAGAGAGAGAG 540
QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGluGlnProVal 200
Db 541 AAACAGAGCCTTCAGACTTCAGAGATCCGAGCCGAGATGTTCCCAAGGGCCGACCGCTG 600
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspHisAspGlnGluGluProAspLeu 220
Db 601 GCGCCCTATTAACACAGCAGGATTCCTCATGATGATACACAGGAGAGAGAGCGGATCTC 660
QY 221 LysThrGluLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240
Db 661 AAACCGGCTGTACTCAAGCGGCGCGCCCAAAATCCAGACAGACACGATGACGAGC 720
QY 241 PheMetGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 260
Db 721 TTCATGGAGAGAGGGGTGAGAGAGATGGGGAGCGGATGGATGGAGGGGAGCGGACG 780
QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
Db 781 GAGTTTCTGAGCGGCGCTCTCGAGACCTGACGAGCGGTACACACGAGGAGACCTGCG 840
QY 281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuLysCysLeuSerArg 300
Db 841 AACATGAGCAGAGAGAGCTCATCAAGAGTACTGGAACGTGAGAAAGTCTCTGCGCG 900
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGluLysAspAla 320
Db 901 ATGAGAGAGAGAGAACACCGCGCTGCGAGAGCAGCGCGCTGGTGGCGAGCAGCGCG 960
QY 321 ArgValArgGluLeuGluLeuLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
Db 961 CGTGTGCGGAGAGCTGAGAGCTGAGCGCTGCGCGCGAGAACTCCAGCTGCGTG 1020
QY 341 ThrGluAsnGluLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
Db 1021 ACCGAGAACGAACTGCACCGCGAGCAGAGAGGAGCGCGCTTTCAGAGTTTGAGAGC 1077

RESULT 3
US-09-983-965-4895
; Sequence 4895, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965

; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4895
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 24-LIB34-008-Q1-E1-F7
US-09-983-965-4895

Alignment Scores:
Pred. No.: 1,11e-55 Length: 414
Score: 691.00 Matches: 130
Percent Similarity: 95.62% Conservative: 1
Best Local Similarity: 94.89% Mismatches: 6
Query Match: 36,18% Indels: 0
DB: 10 Gaps: 0

US-09-972-758a-2 (1-359) x US-09-983-965-4895 (1-414)

QY 155 ArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGlu 174
Db 2 AGACGGCCCTCCAGAGAGAGCGCTTGGAAACCGTACTATACCTGACCTGGAGAGAG 61
QY 175 LysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPhe 194
Db 62 AAGAAAAGTTGATGAGAAAACAGAGCCTGCGAGCTTCGAGAGATTGAGCGGAGAGTTTC 121
QY 195 AlaLysGluGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspHisAsp 214
Db 122 GCCAAGGGCGAGCGAGTTGCTCTCATTAACACAGCAGCGAGTTCCTCATGATGACACGAC 181
QY 215 GlnGluGluProAspLeuLysThrGluLeuTyrSerLysArgAlaAlaLysSerAsp 234
Db 182 CAGGAGGAGCGCGATCTTAATAACCGCGCTCTATCCCAACAGCGCGCTGCCAAATCCGAC 241
QY 235 AspThrSerAspAspPheMetGluGluGluGluGluGluGluGluGluGluGluGluGlu 254
Db 242 GACACCAGCAGTACGAGACTTTATGAGAGACCGGCGAGAGATGGGGGAGCGGAGCGG 301
QY 255 MetGluLysAspGluSerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr 274
Db 302 ATGGAGAGAGAGGAGGAGCTTCTGCGAGGGAGCTTCTGCGAGAACTATGAGCGGTAC 361
QY 275 HisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyr 291
Db 362 CACGCGAGAGCGCTGCAAGACATGAGCAGCAGAGAGCTCATCAAGAGTAC 412

RESULT 4
US-09-954-531-813
; Sequence 813, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22

```

: NUMBER OF SEQ ID NOS: 1392
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 813
: LENGTH: 461
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(461)
: OTHER INFORMATION: n=a,t,g or c
US-09-954-531-813

Alignment Scores:
Pred. No.: 6,81e-50 Length: 461
Score: 629.00 Matches: 120
Percent Similarity: 93.85% Conservative: 2
Best Local Similarity: 92.31% Mismatches: 8
Query Match: 32.93% Indels: 1
DB: 9 Gaps: 0

US-09-972-758a-2 (1-359) x US-09-954-531-813 (1-461)
QY 1 MetAlaGluProHeuSerGluYrGlnHisGlnProGlnThrSerAsnCysThrGly 20
Db 72 ATGGCCGAGCATTTCTTGACAGATATCACACACGACCTCAAACTAGCACTGTACAGGT 131
QY 21 AlaAlaAlaValGlnGlnGlnLeuAsnProGlnArgProGlnAlaGlnGlnArgVal 40
Db 132 GCTGCTGCTGCTCCAGAGAGAGCTGAACCTGAGCGGCCCCAGGCNCACGGAGCGGGTG 191
QY 41 ProGlnGlnAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyArgProGly 60
Db 192 CCCGAGAGAGACAGTAgTGGTCATGAGAGCCTTCCCGCAGTTGGTGGCCTCCGGGG 251
QY 61 ProGlnGlnGlyGlnGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
Db 252 CCGGACGCGGAGAGGAGGAGCTGGAATCCCAACCACTCCCTTGCAGACCCAGGCGCTGTCCA 311
QY 81 GluSerSerCysLeuArgGlnGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
Db 312 GAATCTAGCTGCTCCGAGAGAGGCGGAGAGGCGGAGAGGAGAGGAGAGGAGAGGAGAGG 371
QY 101 GlyAspPheProProProAlaGlnValGlnProThrProGlnAlaGlnLeuLeuAlaGln 120
Db 372 GC-GACTTCCCGGCGGAGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 430
QY 121 ProCysHisAspSerGlnAlaSerGlyLeu 130
Db 431 CCTGTGATGACTCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 460

RESULT 5
US-09-918-995-403/c
: Sequence 403, Application US/09918995
: Publication No. US20030073623A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
: TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
: FILE REFERENCE: 20411-756
: CURRENT APPLICATION NUMBER: US/09/918,995
: CURRENT FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: US/09/235,076
: PRIOR FILING DATE: 1999-01-20
: NUMBER OF SEQ ID NOS: 38054
: SOFTWARE: PastSeq for Windows Version 3.0
: SEQ ID NO 403
: LENGTH: 495
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(495)
: OTHER INFORMATION: n = A,T,C or G

```

```

US-09-918-995-403
Alignment Scores:
Pred. No.: 6,21e-49 Length: 495
Score: 619.00 Matches: 121
Percent Similarity: 98.37% Conservative: 0
Best Local Similarity: 98.37% Mismatches: 2
Query Match: 32.41% Indels: 0
DB: 9 Gaps: 0

US-09-972-758a-2 (1-359) x US-09-918-995-403 (1-495)
QY 237 SerAspAspAspPheMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 256
Db 495 AGCCATGACGACTTTCATGGAAGAGGGGTGAGAGCATGAGGAGACCGTGGATGGCA 436
QY 257 GlyAspGlnSerGlnPheLeuGlnArgAspPheSerGlnThrGlyGlnGlnGlnGlnGln 276
Db 435 GGGCAGCGGACGAGGATTCTTCACAGCGGACTTCGAGAGACGTACGAGCGGTACACACG 376
QY 277 GluSerLeuGlnAsnMetSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 296
Db 375 GAGAGCCTGCGAGAACATGACAGACGAGAGCTCATCAAGAGTACCTGGAACCTGGAGAG 316
QY 297 CysLeuSerArgMetGlnAspGlnAsnAsnArgLeuArgLeuGlnSerGlyArgLeuGly 316
Db 315 TGCCCTGCGCCATGAGAGAGGAGACACCGGCTGGCGCTGGAGACCAAGCGGCTGGT 256
QY 317 GlyAspAspAlaArgValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 336
Db 255 GCGCAGCGAGCGCGCGCTGCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGCGGCGGAGAG 196
QY 337 LeuGlnLeuLeuThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 356
Db 195 CTCGAGCTGCTGACCGGAGAGCAAGCAACTGCAACCGGAGAGGAGGAGGAGGAGGAGG 136
QY 357 PheGlyAsp 359
Db 135 TTTGGAGAGC 127

RESULT 6
US-09-991-936-1712
: Sequence 1712, Application US/09991936
: Publication No. US20030073627A1
: GENERAL INFORMATION:
: APPLICANT: Brandt, Kevin S.
: APPLICANT: Gaines, Patrick J.
: APPLICANT: Stinchcomb, Dan T.
: APPLICANT: Wisniewski, Nancy
: TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
: FILE REFERENCE: FC-6-Cl
: CURRENT APPLICATION NUMBER: US/09/991,936
: CURRENT FILING DATE: 2001-11-21
: PRIOR APPLICATION NUMBER: US/09/543,668
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: 60/128,704
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 1959
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1712
: LENGTH: 549
: TYPE: DNA
: ORGANISM: Ctenocephalides felis
: OTHER INFORMATION:
US-09-991-936-1712

Alignment Scores:
Pred. No.: 1,23e-08 Length: 549
Score: 184.00 Matches: 66
Percent Similarity: 42.29% Conservative: 30
Best Local Similarity: 29.07% Mismatches: 53
Query Match: 9.63% Indels: 78
DB: 9 Gaps: 12

```

US-09-972-758A-2 (1-359) x US-09-991-936-1712 (1-549)

```

QY      34  ProGluAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe--- 52
Db      33  CCACGCGGTGTAAGAAATAGTT-----AGTAAATATCATCAAAAGCTTTTATTA 80
QY      53  ---ProGlnLeuGlyArgProGlyProGluGlyGluGlySerLeuGluSerGlnPro 71
Db      81  TTAATACTGTTGCAACATCAAAATAGTGTAAATTTGAA---AATATTGAATCGGAAT 137
QY      72  Pro-----ProLeuGlnThrGlnAla 78
Db      138  CCAAAATAATATACAAACCCAGTACAGCGCTCCAAATGTCTGCCCGCTTCAACAAAGCGCT 197
QY      79  CysProGluSerSerCysLeuArgGluGlyGlyGlnAsnGlyAspAspSerSer 98
Db      198  -----ACTGAAGCGCCCTCTATCAGAGAGAAAAGGAATGATAGTGAT----- 239
QY      99  AlaGlyLysAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeu 118
Db      239  ----- 239
QY      119  AlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlu 138
Db      240  -----AGAAAGAGAACCAACCAAGTGGT-----GCACAAATATGGGAAG 278
QY      139  GluGluTrpGlyGlnGlnGlnArgGlnLeuGlnLeuGlyLysLysHisArgArgProSer 158
Db      279  -----AAGAAAAAACATAGACGAGCTAAATGT 305
QY      159  LysLysLysArgHisTrpLysProGlyTrpLysLeuThrTrpGluGluLysLysPhe 178
Db      306  AAGAGAAAACCTAAT---AAACCGTAGACAATAAACAGCGTGGACCCACAGTGAAGATGTC 362
QY      179  AspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGln 198
Db      363  ---CAAAAGAAATGATGAGCTGTACTGCCCGTCAATTAATATTCACCTATGGGGCAC 419
QY      199  ProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluPro 218
Db      420  ACATTAGTCCCTTTGTAACACGAACAATTTCTCATGGAAGATCATGAT----- 467
QY      219  AspLeuLysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThr----- 236
Db      468  -----GTCCCTCAACCAAGGATTCATCTGCAGACTCGGACTCTACTAGTGT 512
QY      237  ---SerAspAspAspPheMet 242
Db      513  CGTTCTGGAAGACACTCTNATG 533

RESULT 7
; US-09-974-300-7891
; Sequence 7891: Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ. ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7891
; LENGTH: 2421
; TYPE: DNA
; ORGANISM: Bacillus clausii

```

```
Db 1179 TTACGGCGAGCTGGCAAAATTTGGATTGGGCATACGCCAGCAAGGCTACTTGAATCAAGA 1238
QY 298 -----LeuSerArgMetGluAspGluAsnAspArgLeuArgLeuGluSerLysArgLeuGlu 316
Db 1239 CCGGATTAACCAAAATTTTCATGCATGACAGAACCGGTCTCCGAGGTTTCAGGAAAAAATTT- 1296
QY 316 yGlyAspAspAlaArg-----ValArgGlu 324
Db 1297 ----GACGAGCGCGCGGCAATCAGTCGATTGTATAGCGCAAGCAAGCAATTGGAGAAG 1352
QY 324 uLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGluLeuLeuThrGlu----- 342
Db 1353 AGCCCAACATCTCGTTGAGAGACAGAGCAAGATTATTCATTTATCCGATGCAGAA 1412
QY 343 -----AsnGluLeuHisArg-----GlnGlnGlu 350
Db 1413 CAACGTGCTGACAGAGTTAGAACACAGCAATTGCTGAAGAACAAACAGCATTTGCGAGAGA 1472
QY 350 uArgAlaProLeuSerLysPheGlyAsp 359
Db 1473 AGTGGCGCGCATTTGCAGAAACGGCGCAAC 1500

RESULT 8
US-09-925-302-47
; Sequence 47, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 3773
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-47

Alignment Scores:
Pred. No.: 0.000261 Length: 3773
Score: 147.50 Matches: 105
Percent Similarity: 34.40% Conservative: 46
Best Local Similarity: 23.92% Mismatches: 140
Query Match: 7.72% Indels: 149
DB: 10 Gaps: 23

US-09-972-758a-2 (1-359) x US-09-925-302-47 (1-3773)
QY 3 GluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGlnAlaAla 22
Db 719 GAACCGGAGAAAGCTGAGGCTGAGCACCAGCAAGCTAA-----GCCAGT 760
QY 23 AlaValGlnGluGluLeuAsnPro-----GluArgPro 33
Db 761 GGAGGAGGAGAAAGAAATTCCTCCGGAGAGCAGCTGAGAGGAGAGGAGGCAAGGCAAGCA 820
QY 34 ProGluValGluGluValAlaProGluGluAspSerArgTrpGlnSerArgAlaLeuPro 53
Db 821 CCTGGAACAGCAATGCGCCACCT-----CCA 847
QY 54 GlnLeuGlyAlaArg-----ProGly----- 60
Db 848 TGGCCAGGTGGCGACATGAAAAAGAGATGAGAGACAGTGTGGGTGCTGGAAGCTGC 907
QY 61 -----ProGluGlyGluGlySerLeuGluSerGlnProProPro 73
```

```
Db 908 TGAGAGGATGAAGAGAGACCTCTCAGAGAGACCTTGAGGG-----CCTGAGCCA 955
QY 74 LeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGluGlyGlnAsn 93
Db 956 -----GCCGACAGAGAGAGAGAGTGGCCGCCCTACAGCAACAGCTGGAGAAACCAAC 1006
QY 94 -----GlyAspAspSerSerAlaGlyGlyAspPheProPro--ProAlaGlu 108
Db 1007 GCGGCTGACAGAGAGCTGAGAGACCTCTGCTGTGACCTGGAGCACACGCGCCAGAGCGC 1066
QY 109 ValGluProThrProGluAlaGlu-----Leu 117
Db 1067 GTGCAACCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1126
QY 118 LeuAlaGlnProCys-----HisAspSerGlu 126
Db 1127 CTTCTCCCAAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
QY 127 AlaSerLysLeuGluAlaProAlaAlaGlyGlyGluGluGlu-----TrpGlu 142
Db 1187 CAGGCTCTGTGCTGCGCCCGCGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246
QY 142 yGlnGlnGlnArgGln-----Le 148
Db 1247 GCGKCTCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306
QY 148 uGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTy 168
Db 1307 GGGCAAGAGTGTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1352
QY 168 rLysLeuThrTrpGlnGluLysLysLysPheAspGluGlnGlnSerLeuArgAlaSer 188
Db 1353 -CACAGAGTGGAGAGAGATGAAGACGACGCTGGAAGAGCTGAGAGAGCA-GCTCCAGGCCA 1410
QY 188 gLLeuArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnP 208
Db 1411 CCGAAGATGCCAACCTCGGTTGGAGCTCAACTGCAGGC-----1450
QY 208 eLeuMetAspAspHisAspGlnGluProAspLeuLysThrGlyLeuTyrSerLysArg 228
Db 1451 -----CATGAAGGCCAGAGTTCAGCGGAGCTTCAGAGGCCGCG 1488
QY 228 gAlaAlaAlaLysSerAspAsp-----ThrSerAspAspPheMetGlu 243
Db 1489 ACCAGCAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1548
QY 243 uGluGlyGlyGluGluAspGlyGlySerAspGlyMetGly-----GlyAs 258
Db 1549 AGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
QY 258 pGlySerGlu-PheLeuGlnArgAspPheSerGlnThrTrpGluArgTyrHisThrGlu 278
Db 1609 TGGACCTGAAGAGACCTGGAAGCG-----CACATGCACT 1641
QY 278 eLeuGlnAsnMetSerLysGlnGluLeuLeuLysGluTyrLeuGluGluGluGluGlu 298
Db 1642 CG---GCCAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1689
QY 298 eLysArgMetGluAspGluAsnAspArgLeuArgLeuGluSerLysArgLeuGlyGly 318
Db 1690 AGGCCCGAGTGAAGAGC-----TCCATGCCGAGCTGGATGATGACCCGCGCTCTGCTG 1743
QY 318 sPAsp-----AlaArgValArgGluGluGluGluGluGluGluGluGluGluGluGlu 336
Db 1744 AGGAGATCTCTGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1803
QY 336 snLeuGlnLeuThrGlnAsnGlnLeuHisArgGlnGlnGluArgAlaPro 353
Db 1804 TGATCCAGTTG-----CAGAGAGAACTGGCAGCCGCGAGATGCCAAGCGCCA 1850

RESULT 9
US-10-037-270-248
; Sequence 248, Application US/10037270
```

OY	11	ThrSeAsnCySThGlValAlaValGlnGlu-----GLeuAsnPProGluArg	32
Dd	3948	ACTACAGSATTACACAGAGCTTTCTTCAGGAGGAGACC GCCCAAACTAAACCTTAGCGC	4007
OY	33	ProPProGlyAlaGluGluArgValProGluGluAspSerArgTrpIleSerArgAlaPhe	52
Dd	4008	TGGATCCGGACGTCTGAGAGGAGAAGAACACTTTC A-----	4046
OY	53	ProGlnLeuGlyGlyArgProGlyProGluGlyGlyGlySerLeuGluSerGlnProPro	72
Dd	4047	GGACGACGAGGAGGAGGAGGAGGAGGCCACCAAGAACTGGAGAGCAAGTGCTGGCCT	
OY	73	ProLeuGlnThrGlnAlaCysProGluSerSerCysteLeuArgGluLy-----Glu	89
Dd	4107	GCAGTCGCCAGTTGGCTGATACCAAGAGAAAGATGACGACGCTGGGMAAATTGAAG	4166
OY	90	LysGlyGlnAsnGlyAspAspSerSerAlaGlyAspPheProProAlaGluVal	109
Dd	4167	TCTGGAAAGACCAAGAGAAAGCTTCTGAAAGACCGGAGGGCCCTGAGCCACGC-----	4220
OY	110	GluProThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLys	129
Dd	4221	--CCTGGAGAGAAAGCCACTGGCTGATGACAACACTGGAGAGACCAGCAAGCCGCTGCA	4277

OY	130	LeuGIValA-----	-----ProIAlAlaGIValGIValuGIuTTP	141
Db	4278	GCAGGAGCTGGACGACCTCAGGTGGACCTGGACACCAACGACGCGCCACAGTTCGCCCAACTT	-----	4333
OY	142	GIYGIInGIInIArgIn-----	-----	147
		: :	: :	
Db	4338	GGAGAGAGCAAGAAAGATTGGACCCACTGTATGACAGAAAGAAAGACATCTCTGCTC	-----	4397
OY	148	-----	-----LeuGIlyIyIyIySHIArgArIArgProSerLySIyIyIy	161
Db	4398	GCTATGCCGAGAGACGGGACCCGGCCGAAAGCCGACGACAGAGAAAGAAACCAAAAGCCC	-----	4457
OY	162	ArgHISIRPLySProTyrTyrLySLeuThrTPGILuIyIyLySIyIyPheAsp----	GIu	180
				: : :
Db	4458	TGTCACTGGCCCGGGCCCTCGAGAGAACCCCTGGAGGCC-AAAGAGAGATTGAGAGGCGAG	-----	4516
OY	181	LySGLISerLeuArGIAlSer-----	-----	187
Db	4517	AACCAACAGACTCGAGAGACATGTGAGAGACTCATGAGCTCCAAAGATGATGTGGAAAA	-----	4576
OY	188	-----ArgILleArGIaGIuIleIrePheAlIySGILyGIInProValAlaProTyrAsn	204	
Db	4577	AACGTTTACGAACTTGAAAATAATCCAAACGGGCCCTAGAGCAGACAGAGTGTGGAGAAATGAG	-----	4636
OY	205	ThrTrhGIInPheIleuMetAspAspHisAspGIInGIuIuProAspLeuThrGIyIleu	224	
Db	4637	ACCCAGCTGGAGGACTTGGAGAGCAACTTCAGAGCCACAGAAAGATGCCAAACTTCGCTTG	-----	4696
OY	225	TyrSerLySIyArGIAlAlaIySIerAsp-----	-----AspThrSerAspAsp	239
Db	4697	GAGGTCAACATTCAGGCTGAGAGGCGCGATTCGAGAGAGACCTGCACAAACAGGATGAG	-----	4756
OY	240	--AspPheMetGIuGIuGIyGIyGIuGIuAspGIyGIySerAspGIyMetGIyIyIyAspG	259	
Db	4757	CAGATTGAAGACAAAGAACCGCTGCTGATCAACAGAGGCGCGAGCTTCGACGCGGACTTG	-----	4816
OY	259	Iy-SerGIuPheIleuGIInIArgAspPheSerGIuThrTyrGIuArGIyThISthrGIuSer	278	
Db	4817	GAGATGTAGAGAAACACAGCGCGCTTGTCTACTTCAAAAGAAAGATTCGAGATGAC	-----	4876
OY	279	LeuGIIn-----	-----AsnMetSerLySGInGIuLeuIleIyS	289
Db	4877	CTGAAGGACCTTCGAAGCCCAATATCGAGGCTGGCAACAAACCTCGGATGTGAGGATTAAG	-----	4936
OY	290	GIuTyrGIeugIuIeugIuIyCySIuSerArgMetGIuAspGIuAsnAsnArgIeueArg	309	
Db	4937	-----CAGCTTCGGAAGCTTCAGGCTCAGTGTGAAGATTACCAACAGT-----	-----GAA	4981
OY	310	LeuGIuSerLySIyArgIeugIyGIyAspAsp-----	-----AlaArgValArgGIuIeugIuLeu	327
Db	4982	TTAGAAAGAGCTGCTGCATCCAGAGATGAATTTTCTCAATCCAAAGAGAGCTGAAGAAG	-----	5041
OY	328	GIuIeueAspArgIeueArGIaGIuAsnLeuGIuIeueuThrGIuAsnGIuIeue-----	-----	345
Db	5042	AAATTGAAGAGCTCTGGAGACGAGAAATCTTCAATTG-----	-----CAGAGAGAACTTGCTCA	5095
OY	346	-----	-----HisArgGIInGIuIuArgAlaProIeueSer	355
Db	5096	TCTGAGCGAGCCGCGCAGACACGCGGACAGAGAGAGATGAGCTTGGCG	-----	5143

; PRIOR APPLICATION NUMBER: US 60/292,336
 ; PRIOR FILING DATE: 2001-05-22
 ; PRIOR APPLICATION NUMBER: US 60/295,798
 ; PRIOR FILING DATE: 2001-06-06
 ; PRIOR APPLICATION NUMBER: US 60/297,457
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,884
 ; PRIOR FILING DATE: 2001-06-19
 ; PRIOR APPLICATION NUMBER: US 60/303,459
 ; PRIOR FILING DATE: 2001-07-09
 ; NUMBER OF SEQ ID NOS: 1740
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 1666
 ; LENGTH: 4301
 ; TYPE: DNA
 ; ORGANISM: Rattus norvegicus
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_019206
 US-09-917-800A-1666

Alignment Scores:
 Pred. No.: 0.000516 Length: 4301
 Score: 145.00 Matches: 96
 Percent Similarity: 37.94% Conservative: 55
 Best Local Similarity: 24.12% Mismatches: 160
 Query Match: 7.59% Indels: 87
 DB: Gaps: 20

US-09-972-758a-2 (1-359) x US-09-917-800A-1666 (1-4301)
 QY 2 AlaLupProheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAla 21
 Db 1028 TCAGAACCAAGACGACGCTGAG-----CACCAAGACGACGCTGACCTGGAGCAGATGCA 1081
 QY 22 AlaAlaVal-----GlnGlnGlnLeu-----AsnProGluArg 32
 Db 1082 CAGACGATTTCAGCAAGAAATCAACGCCAAGAAATTTATGACAGCTGGAGCTAAGAA 1141
 QY 33 ProProGluLysAlaGluGluArgValProGlnGluAspSerArgTrpGlnSerArgAlaPhe 52
 Db 1142 CCTGGAGCGCGCAGCAGCAGAGAGTGGAGAGATGGACAGCAGCAGCTGCG--- 1198
 QY 53 ProGlnLeuGlyLysArgProGluGlyGlnGlySerLeuGlnSerGlnProPro 72
 Db 1199 TCCGACAGAGAGAGCCAGCGGATCCGCTCGAGGACAGATGAGACTACGCCAGGTCCA 1258
 QY 73 ProLeuGlnThrGlnAlaLysProGluSerSerCysLeuArgGlnGlyGluLysGln 92
 Db 1259 -----AGAGCAGCTCAAGCAGATGAA 1279
 QY 93 AsnGlyAspAspSerSerAlaGlyLysAspPheProProAlaGluValGluProThr 112
 Db 1280 GAAGGAGGTGAAGATGAGGTTGAGAACTGCCCGCAACGCGAAGAGAGATGAA 1339
 QY 113 ProGlu-----AlaGlnLeuLeuAlaGlnPro-----CysHisAsp 124
 Db 1340 GCAGAGATGAGAGCAGCAGCAGCAAGAAACACTGCTGGACCCAGACTTTGTAGCCA 1399
 QY 125 SerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlnGlnTrpLysGlnGln 144
 Db 1400 GCAGAA-----GGAAGACCTTGAGCTGGCTGACAGAGAGCTCAGCGCAGAAA 1447
 QY 145 Gln-ArgGlnLeuGlyLysLysHisArgArgArgProSerLysLysLysArgHisTr 164
 Db 1448 CAGCGGTGATCTGTGACAAAGAA-----CGTATTGCCCTTACCAAGAAAGAGAGACTCTCT 1504
 QY 164 PLSProGlyTrpLysLysLeuThrTrpGlu-----GluLysLysLysPheAspGluLysGln 183
 Db 1505 CCGAGACCGAGAGCGCCCTGTGGAGATGAGAGACACAGCTTACAGAGAGAGATCA 1564
 QY 183 rLeuArgAlaSerArgGlnArgAlaGlnLeuPheAlaLysGlnProValAlaProTy 203
 Db 1565 GCTGTGAAGCAGCAGCTTAAGGACCACTACTCTGAGCGGCGCATGAC----- 1613

QY 203 rAsnThrThrGlnPheLeuMetAspHisAspGlnGlnGluProAspLeuLysThrGln 223
 Db 1614 -----CTGCTGGCAAGCCAGGAAGAGCGGAGCAGATGCACCGC-- 1655
 QY 223 yLeuTyrSerLysArg-----AlaAlaAlaLysSerAspAspThrSerAspAspPh 241
 Db 1656 -----TACAAACACCGTATGATGAGCAGCTGAAGGTCAAGACAGCAGAGAGAGCGCG 1711
 QY 241 eMetGlnGlnGlyLysGlnGluAspGly-----GlySer 253
 Db 1712 GCTAACCAAGATCCAGAGAGAGAGCGGCAAGACCCGATGCCATGTACAAAGAGCTT 1771
 QY 253 pGlyMetGlyGlyAspGlySerGlnPheLeuGlnArgAsp-----PheSerGln 269
 Db 1772 GCACATCAATGTGTGGCGGAGTGTCCCGACGAGCGGCGAAGAGTCAACAGCTTCCCA 1831
 QY 269 uThrTyrGlu-----ArgTyrHisThrGlnSerLeuGlnAsnMetSerLysGlnGlu----- 286
 Db 1832 GCAGGAAGAGAGAGAGCAGAGCGGAGAGCTGACAGCAGCAGAGAAACACAGAGACCA 1891
 QY 287 -----LeuLys-----GluTyrLeuGlnLe 294
 Db 1892 GATGCGAGACATGTGTGCGACAGTCCGAGAGCAACATGAACGAGCTGCAGACAGCTCAGAA 1951
 QY 294 uGluLysCysLeuSerArgMetGlnAspGlnAsnAsnArgLeuArgLeuGlnSerLysAr 314
 Db 1952 TGAAGAGTTCATCTGTTAGTGACATGAACCCAGAGAGCTGAAG----- 1997
 QY 314 gLeuGlyGlyAspAspAlaArgValArgGlnLeuGlnGluLeuAspArgLeuArgAl 334
 Db 1998 -----GCCCTGAGAGAGAGCCATACCAAGAGCTTGAAGGATGGCAGACAGCTTCGGCC 2053
 QY 334 aGluAsnLeuGlnLeuLeuThrGlu-----AsnGlnLeuHisArgGlnGlnGln 350
 Db 2054 ACCCAAAAAGGCCCTGGAAGAGATTGAACCAAGAAAGCGGCAAGCA 2105

RESULT 12
 US-09-291-417-106
 ; Sequence 106 Application US/09291417A
 ; Publication No. US20030050230A1
 ; GENERAL INFORMATION:
 ; APPLICANT: PLOMAN, GREGORY
 ; APPLICANT: MARTINEZ, RICARDO
 ; APPLICANT: WHITE, DAVID
 ; TITLE OF INVENTION: STEP20-RELATED PROTEIN KINASES
 ; FILE REFERENCE: 240/300
 ; CURRENT APPLICATION NUMBER: US/09/291,417A
 ; CURRENT FILING DATE: 1999-04-13
 ; EARLIER APPLICATION NUMBER: US 60/081,784
 ; EARLIER FILING DATE: 1998-04-14
 ; NUMBER OF SEQ ID NOS: 147
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO: 106
 ; LENGTH: 2962
 ; TYPE: DNA
 ; ORGANISM: Full Length Mammalian (Human) GEX2
 US-09-291-417-106

Alignment Scores:
 Pred. No.: 0.000792 Length: 2962
 Score: 141.00 Matches: 103
 Percent Similarity: 39.65% Conservative: 56
 Best Local Similarity: 25.69% Mismatches: 145
 Query Match: 7.38% Indels: 98
 DB: Gaps: 23

US-09-972-758a-2 (1-359) x US-09-291-417-106 (1-2962)
 QY 8 GlnTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAla----- 23
 Db 1784 GAGCATGGAAGCAAGACCAAGCTGAGTACAA-GCATGAGCTGCACTGGAGCAAAATGCA 1842

```
QY 24 -----ValGInGluGluLeu-----AsnProGluArgPro 33
DB 1843 TAAAGCTTTTGAACAGAAATACAGCCAGAGAAAGTTCTTGTGACAGGAATTAGACAA 1902
QY 34 ProGlyAlaGluArgValProGlu---GluAspSerArgTyrGlnSerArgAlaPhe 52
DB 1903 CTTGGAGCGCTGACAGAAAGCAGCAAGTGGAGCAAGCATGCGGTGCG--- 1959
QY 53 ProGlnLeuGlyArgProGlyProGlnGlyGlySer---LeuGlnSerGlnPro 71
DB 1960 CCGCCGGAGAGAGCGGCGGATCGCTGGAGCAGGATCGGAGCTACACAGGTCCA 2019
QY 72 ProProLeuGlnThrGlnAlaGlyCysProGlnSerCysLeuArgGluGlyGly 91
DB 2020 AGAGCAGCTCAAACT-----GATCAAGAAAGAGGT 2049
QY 92 GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaGluValGluPro 111
DB 2050 GAA-----GACGAGGTGGAGAGCTCCCGCCAGCA---GCGGAGGA 2091
QY 112 ThrProGluAlaGlu-----LeuLeuAlaGlnPro----- 121
DB 2092 AAGCATGAAACAGACAGATGGAGAGCAGCAGAAAAGACACTTCTGACCGGAGCTT 2151
QY 122 CysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluTyr 141
DB 2152 TGTAGCCACACAGAA-----GGAGACCTGGAGCTGGCATGAGAGGCTCAC 2199
QY 142 GlyGlnGlnGln-ArgGlnLeuGlyLysLysHisArgArgArgProSerLysLys 161
DB 2200 CACCGACACACAGCGGAGATCTGTGACAGAG---CGGAGTGGCTCATGAAGAAGA 2256
QY 161 SarGlnHisTyrLysProTyrTyrLysLeuThrTyrGlu---GluLysLysLysPhe 180
DB 2257 GAGACTCTCTTCGAGACCGGAGACAGCCCTGTGGAGATGAGAGAGCAGCTGAGCA 2316
QY 180 uLysGlnSerLeuArgLaserArgLLeuArgAlaGlnMetPheAlaLysGlyLlnPro 200
DB 2317 GAGGACACAGCTGTGAGACAGCAGCTCAAGACAGCACTTCTTCCAGCGGACAG 2374
QY 200 lAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAsp 220
DB 2375 -----CTGCTGCCAAGCATGAGAGAGCGGAGCAGCAT 2409
QY 220 uLysThrGlyLeuTyrSerLysArg-----AlaAlaAlaLysSerAspSerPhe 238
DB 2410 GCAGCGC-----TACACACAGCGCATGATAGACACCTGAAAGTGGCGGACAGCA 2463
QY 238 PASpAspPheMetGluGluGlyGlyGluGluAspGly----- 250
DB 2464 AAAGCGCGGCTGCCCAAGATCCAGAGAGTGAAGGAGCAGCGCATGGCATGTACAA 2523
QY 251 -GlySerAspGlyMetGlyLysAspGlySerGluPheLeuGlnArgAsp----- 266
DB 2524 GAAGAGCTCCACATCAACCGCGGCGGAGCAGCTGAGAGGCTGAAGAATCAAGCA 2583
QY 267 -PheSerGlnThrTyrGlu---ArgTyrHisThrGlnSerLeuGlnAsnMetSerLys 285
DB 2584 GTTCTTCCAGCAGAGAGAGAGAGAGAGAGTCCGAGCGGTGACGACAGCAGGAACA 2643
QY 285 nGlu-----LeuLeuLys-----GluTyr 291
DB 2644 CGAGAAACAGATGGCGGACATGGCGGATGGCGAGCAACATGACGAGCTGACGA 2703
QY 291 rLeuGlnLeuGluLysCysLeuSerArgMetGluAspGlnLysAsnThrArgLeuArg 311
DB 2704 GCTGCAAGATGAAAGGTCCACCTCTGTGAGACAGCAAAACCCAGAAACTGGAAG 2758
QY 311 uSerLysArgLeuGlyLysAspAspAlaArgValArgGluLeuGlnLeuGlnLeuAsp 331
DB 2759 -----GCCCTGGATGAGAGCCATACCAAGAACTGAAGAGATGGCGGACAA 2805
QY 331 gLeuArgAlaGluAsnLeuGlnLeuLeuThrGlu---AsnGluLeuHisArgGlnGln 350
```

```
DB 2806 GCTTCGGCCCGCCGACAGAAAGCTCTGAGAGAGGATCTGACACAGAAAGCGGAGAGGA 2865
QY 350 u 350
DB 2866 G 2866

RESULT 13
US-10-175-523-186
; Sequence 186, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Paley, Michael
; APPLICANT: Rajan, Pithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 186
; LENGTH: 6537
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-175-523-186

Alignment Scores:
Pred. No.: 0.00298 Length: 6537
Score: 139.00 Matches: 88
Percent Similarity: 34.07% Conservative: 66
Best Local Similarity: 19.47% Mismatches: 147
Query Match: 7.28% Indels: 152
DB: 9 Gaps: 16

US-09-972-758a-2 (1-359) x US-10-175-523-186 (1-6537)
QY 7 serGluTyrGlnHisGlnProGlnThrSerAsnGlyThrGly----- 20
DB 3442 AATAAATATTTTGCATTTCCTCCCGGAGACCGCCAGTCACAGGGGAGAGGAGCTGCTG 3501
QY 21 -----AlaAlaAlaValGlnGluGluLeuAsnProGluArgProProGlyAla 36
DB 3502 GAGCTGTCTAAGACAGACAGTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAA 3561
QY 37 GlnGluArgValProGluGluAspSerArgTyrGlnSerArgAlaPheProGlnLeu 56
DB 3562 GAGGAGGAAGAAAGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAA 3615
QY 57 GlyArgProGlyProGluGlyGlySerLeuGlnSerGlnProProProLeuGlnThr 76
DB 3616 -----GAGAGAGAGAAATATTTCAAGCTCTCCCAAGATTGACGAAA 3660
QY 77 GlnAlaCysProGlnSerSerCysLeuArgGlu----- 87
DB 3661 -----CCACAGTCAGTTGCCATTAAGAGAAAGAGGCGTTTGTACTAAAGAGAAA 3711
```

87 ----- 87
Db 3712 AGGGGCTGTAACGAGGAGATCAACAGCATGTAAACAACAGACCATTTCCAGAGC 3771
Qy 88 -----GlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyAspPhe 103
Db 3772 ACAGAACTAGTGAAGCCCTTGGACAACTGACATGAGAGAGG----- 3816
Qy 104 ProProProAlaGluValGluProThrProGluAlaGluLeu----- 117
Db 3817 ---CCATGCCACAGCTGAGCCTACCTGTGAGATTGAACTGGAGAGATGGCAGAG 3873
Qy 117 ----- 117
Db 3874 CCACTCTGAGAAAGCATTCACGATCAGCTGGGAGAAACACAAACAGAGAGAG 3933
Qy 118 -----LeuAlaGlnProCysHisAspSer----- 125
Db 3934 GAAGAGAAAGACAATCATCTGCTTCAAGATGCTGACCTTGTAGAAACAATATGATGAT 3993
Qy 126 -----GluAlaSerLysLeuGluValProAlaAlaGlyGlyGlu 139
Db 3994 GATTCAAGTACTTTGAAAGAGGCGAGTAAAGACAAATCCGACCTTAAAGTGCACAA 4053
Qy 140 GluThr-GlyGlnGlnAlaArgGlnLeu-----GlyLysLysHisArgArg 156
Db 4054 GTGTGGCCAAAGAGAACAAAGCCGGCTCTATCTAGTGGAGGCAAAACAAAGAGAGAG 4113
Qy 156 gProSer-----LysLysLysArgHisTrpLysProTyrTyr 169
Db 4114 ACCGATTTAACTGAATTTGTACACCCCGCCAGAAACACCCATGGAG--CCTGACGAGCA 4172
Qy 169 sleuthThrProGluLysLysLysPheAspGlu---LysGlnSerLeuArgAlaSer 188
Db 4173 GGTAACGTGAAGACAGACGAGACTTCAGAGAAACCAAGCCAGTCCCATCAG 4232
Qy 188 gIleArgAlaGluMetPheAlaLysGlyGln----- 198
Db 4233 GATTGAGAGAGAGTCAAGAACTGGGCAACCCCTTGTCCCTCAAGAGCAAAACAGAG 4292
Qy 199 -----ProValAlaProTyrAsnThrThrGlnPheLeuMetAspAsp 213
Db 4293 GGAAGAACATGTGCCCTGTAACTCA---AACACATCA-----CCAG 4334
Qy 213 sAspGlnGluGluProAspLeuThrGlyLeuTyrSerLysArgAlaAlaLysSe 233
Db 4335 TGAATAACCAAGATGATCTC-----ATCAACCTGAGGAGAGAGCAAGAGAGGA 4385
Qy 233 rAspAspThrSerAspAspPheMetGlnGlyGlyGlnLysAspGlyGlySer 253
Db 4386 GGAAGAGAGAGAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4440
Qy 253 pGlyMetGlyLysGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGlu 273
Db 4441 -----GGAATGTAGAAATAAGATCCAGATGCTGCTTAAAGCCAAAGAAA 4484
Qy 273 gTyrHisThrGluSerLeuLysMetSerLysGlnGluLeuLysGluTyrGlu 293
Db 4485 AGAGAGAACCAAGAAATCTCCACGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4544
Qy 293 uLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGluSer 313
Db 4545 GGAAGAGAGAA-----GAGCATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4589
Qy 313 sArgLeuGlyLysAspAlaArgValArgGlnLeuGlnLeuGlnLeuAspArgLeu 333
Db 4590 TGCCGATGAG 4649
Qy 333 gAlaGluAsnLeuGlnLeuLeuThrGluAsnGlu 344
Db 4650 CAGAGAAAGCTTCAAG 4683

US-09-927-597-1
; Sequence 1, Application US/09927597
; Publication NO. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Fady
; APPLICANT: Berlaud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CYTOPO18
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ. ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 1
; LENGTH: 5835
; TYPE: DNA
; ORGANISM: Human
US-09-927-597-1
Alignment Scores:
Pred. No.: 0.00291 Length: 5835
Score: 138.50 Matches: 123
Percent Similarity: 33.51% Conservative: 64
Best Local Similarity: 22.04% Mismatches: 130
Query Match: 7.25% Indels: 245
DB: Gaps: 29
US-09-972-758A-2 (1-359) x US-09-927-597-1 (1-5835)
Qy 14 GlnThrSerLysncyThrGlyAlaAlaValGlnGluLeuAsnProGlu----- 31
Db 1599 CAACCTCCAGAGGTGTGCTGCGCTGCGA---CGAGGAATGCTGTTCCCAAGCCAC 1655
Qy 32 -----ArgProProGlyAlaGlu 38
Db 1656 GGACAACTCTTCTGAGAGAGCTGTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1703
Qy 39 ArgValProGluLysSerArg----- 46
Db 1704 CAAGTTCAG 1763
Qy 47 TrpGln----- 48
Db 1764 TGGGAAGGTGACTATATGCGAGTGTGCTGACCAAGAAATATGACACCCCTGAATGA 1823
Qy 49 SerArgAlaPhePro-----GlnLeuGlyArg----- 58
Db 1824 CAACGTACTTCCCTGCTCAATGCGCTCTCCGACAACTTTGTGGCCGACCTGTGAAGGA 1883
Qy 59 -----ProGlyProGluGly-----GluLysSerLeuGluSerGln 70
Db 1884 CGTGAGCCGATCCTGCGCTGGAGAGAGATGGCAAGATGACGAGAGAGAGAGAGAGAGAG 1943
Qy 71 ProProProLeuGlnThrGlnAlaCys-ProGluSerSerCysLeuArgGlnGlyGly 90
Db 1944 CG-CCCTCAAGACCAAG 2002
Qy 90 sGlyGlnAsnGlyAspAspSer----- 98
Db 2003 TGGCAAGCTGATGACACACCTTACGACACACACACACACACACACACACACACACAC 2062
Qy 98 ----- 98
Db 2063 CCAACACAG 2122
Qy 99 -----AlaGly-----GlyAspPheProProProAlaGlu--- 108
Db 2123 GCAATGGGTGCTGAG 2182
Qy 109 -----ValGlnProThr-----ProGluAlaGluLeuLeuAlaGlnProCysHis 123

```

Db      2183 TCACGAGATTCCGACAGCCTACGAGATCTGCGCGCATCCATCCCAAGGCTTCA 2242
Qy      123 saspsergluaser-----
Db      2243 TCGAGGGGAGACGAGCCCTGCATTCATGATATCAAAAGCCCTGACCTGACCCCAACTTAT 2302
Qy      129 -----LysleuGluAlaProAlaAlaG1 136
Db      2303 ACAGGATAGGCGAGAGCAAAATCTTCTCCGAACTGGCGTCTGCGCCACCTAGAGAGG 2362
Qy      136 Y-----GlyGluGluGluTrp----- 141
Db      2363 AGCGAGATTGAGATCACACCATGTCATCATGCGCTTCGAGCGAGATGTCGTGCTACT 2422
Qy      142 -GlyGluGlu-----GlnArgGlnLeuGlyLysLys 152
Db      2423 TGGCCAGAAAGGCTTTGGCAAGAGGAGCAGCAGCAGCTGACCGCCATGAAGGATTCAGA 2482
Qy      152 shisArgArgArgProSerLysLysArgHisTrp----- 164
Db      2483 GGAACCTGGCGCG-CCTACCTC-AAGCTGCGAAGTGCAGTGTGAGAGGCTTTCCACCA 2540
Qy      165 ----LysProLysLysLysLysLysLysLys-----LysPheAspG1 180
Db      2541 ACTGAAGCCACTGCTGCAGGTGACGCGAGAGAGAGAGATGACGCCAAGAGAGATGA 2600
Qy      180 uLysGlnSerLeuArgAlaSerArg1LeArgAlaGluMetPheAlaLysGlyGlnProVa 200
Db      2601 ACTCGAAGAACCAAGAGAGCGGACAGCAGAGAGCAAG----- 2637
Qy      200 lAlaProLysArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 218
Db      2638 -----AATGAGCTTAAGGAGCTGTAACAGAGAGAGAGAGAGAGAGAGAGAGAA 2687
Qy      218 oAspleu-----LysThrGlyLeuLysLysLys----- 227
Db      2688 GAACCTGCTACAGAGACAGCTGCAGAGCAGAGCAGAGAGCTGTATGAGAGGCTGAGGAGAT 2747
Qy      228 -----ArgAlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGlu 245
Db      2748 GCGGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2789
Qy      245 yGlyGluLysProGlySerAspGlyMetGlyLysAspGlySerLysLysLysLysLys 265
Db      2790 -ACTGCATGATGAGGCGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2848
Qy      265 rGAspPheSerGluThrTyrgLysLysLysLysLysLysLysLysLysLysLysLysLys 285
Db      2849 CTGAAGAGAGAGATGCGCCAGCAGATG---CTGAGACCTTGAAGAGAGAGAGAGAGAG 2905
Qy      285 lngluLeuIleuGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 300
Db      2906 AGGAAGCTGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2965
Qy      301 -----MetGluAspGluAsnArgLysLysLysLysLysLysLysLysLysLys 313
Db      2966 AACTGAGAGATGAGATCTGCTCATGATGATTCGAAACAATTAACATAAAGAGAGAGAG 3025
Qy      313 ysArgLeuGlyLysAspAspAlaArgValArgLysLysLysLysLysLysLysLysLys 331
Db      3026 AACTGCTT-----GAGGAGAGAGATTAAGTACTTAAGCAGCAAAATCTTCCAGAGAGG 3076
Qy      331 rGLeuAlaGluAsnLeu-----GlnLeuLeuThrGlu 343
Db      3077 AAGAAAGGCCAAGATCTTAACCAAGCTGAAGAAACAGAGATGATCATGATTTCAACAC 3136
Qy      343 snGluLeuHisArgGlnGlu-----ArgAlaProLeuSerLys 356
Db      3137 TGGAAATGGCGCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3186

```

RESULT 15
US-09-927-597-3

```

: Sequence 3, Application US/09927597
: Publication No. US20030032018A1
: GENERAL INFORMATION:
: APPLICANT: Malik, Fady
: APPLICANT: Betaud, Christophe
: APPLICANT: Freedman, Richard
: APPLICANT: Craven, Andrew
: APPLICANT: Sakowicz, Roman
: APPLICANT: Hartman, James
: TITLE OF INVENTION: Human smooth muscle myosin heavy chain
: FILE REFERENCE: CYT0018
: CURRENT APPLICATION NUMBER: US/09/927,597
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 3
: LENGTH: 5937
: TYPE: DNA
: ORGANISM: Human
: US-09-927-597-3

Alignment Scores:
Pred. No.: 0.00297 Length: 5937
Score: 138.50 Matches: 123
Percent Similarity: 33.51% Conservative: 64
Best Local Similarity: 22.04% Mismatches: 130
Query Match: 7.25% Indels: 245
DB: 9 Gaps: 29

US-09-972-758a-2 (1-359) x US-09-927-597-3 (1-5937)

Qy      14 GlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGlu----- 31
Db      1599 CAACCTCCAGAGTGCTGCTGCGCGCTGCGA---CGAGAAATGCTGTGTTCCCAAGGAC 1655
Qy      32 -----ArgProGlyAlaGluGlu 38
Db      1656 GGACAAGCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1703
Qy      39 ArgValProGluAspSerArg----- 46
Db      1704 CAAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1763
Qy      47 TrpGln----- 48
Db      1764 TGGGAAGTGAATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
Qy      49 SerArgAlaPhePro-----GlnLeuGlyArg----- 58
Db      1824 CAACGTGACTTCTGCTCAATGCTCTCCGACACAGTTGTGTGCGCAGCTGTGAGAGA 1883
Qy      59 -----ProGlyProGluGly-----GlnLysSerLeuGluSerGln 70
Db      1884 CGTGAGACGAGATGCTGAGGCTGAGACGAGATGCGCAAGATACAGAGAGAGAGAGAGAG 1943
Qy      71 ProProLeuGlnThrGlnAlaCys-ProGluSerSerCysLeuArgGluGlyGlu 90
Db      1944 CG-CTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2002
Qy      90 sGlyGlnAsnGlyAspAspSer----- 98
Db      2003 TGGGCAAGCTGATGACACGCTACGACACACACAGCCCAACTTGTGCGCTGCATCATCC 2062
Qy      98 ----- 98
Db      2063 CCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2122
Qy      99 -----AlaGly-----GlyAspPheProProAlaGlu-- 108
Db      2123 GCATGGGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2182
Qy      109 -----ValGluProThr-----ProGluAlaGluLeuLeuAlaGluProCysH 123

```

```
Db      2183  TCCAGAGATTCCGCCAACGCTACGAGATTCCTGGCGGCGAATGCCATCCCAAGGCTTCA 2242
QY      123  sAspSerGluAlaSer----- 128
Db      2243  TGGACGGGAAGCAGAGCGCTGCTATTCATGATCAAAAGCCCTGGAACCTTGACCCCACTTAT 2302
QY      129  -----LysLeuGlyValProAlaIaG1 136
Db      2303  ACAGGATAGGCGACAGCAAAATCTTCTCCGAACGTGGCTCCCTGCCCCACCTAGAGGAGG 2362
QY      136  y-----GlyGluGluIuTrp----- 141
Db      2363  AGCGAGATTGAAGATCACCAGATGTCAATACATGCGCTTCACAGCGCATGTGTGCTACT 2422
QY      142  -GlyGlnGln-----GlnArgGluLeuGlyLysLysLys 152
Db      2423  TGGCCAGAAAGGCTTTTGGCAGAGGACAGCAGCTGACCGCCATGAGGTGATTCAGA 2482
QY      152  sHisArgArgProSerLysLysLysArgHisTrp----- 164
Db      2483  GGAACCTGCGCCG-CCTACCTC-AAAGCTGCGGAACCTGGCAGCTGTGAGAGCCTTTTCACCA 2540
QY      165  ---LysProTyrTyrLysLeuThrTrpGluGluLysLys-----LysPheAspG1 180
Db      2541  AGTGAAGCCACTGCTGCGAGCTGACACGAGCAGAGAGAGATGCAGGCCAAGAGAGATGA 2600
QY      180  uLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVa 200
Db      2601  ACTGCAGACACCAAGAGCGCGCAGCAAGAGCAGAG----- 2637
QY      200  lAlaProTyrAsnThrGlnPheLeuMetAspAspHisAspGln-----GluGluPr 218
Db      2638  -----AATGAGCTTAAGAGACTGTGACAGACAGAGACACTGCGACCTGACCGAGAGAA 2687
QY      218  oAspLeu-----LysThrGlyLeuTyrSerLys----- 227
Db      2688  GAACCTGCTACAGAACAGCTGCAGGCGCAGAGACAGAGCTGTATGTCAGAGCGTGAAGAT 2747
QY      228  -----ArgAlaIaAlaLysSerAspAspThrSerAspAspPheMetGluGluG1 245
Db      2748  CCGGCTGCGGCTGCGGCGCAAGAAACAGAGAC-----CTGAGAGAGAT 2789
QY      245  yGlyGluGluAspGlySerAspGlyMetGlyGly-AspGlySerGluPheLeuGlnA 265
Db      2790  -ACTGCATGACATGGAGGCCCGCCCTGAGAGAGAGAGACAGGCGCCAGCAGCTACAGG 2848
QY      265  rGAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysG 285
Db      2849  CTGAAGAGAGAGATGGCCAGCATG--CTGACCTTGAAAGACAGCTGGAGGAGG 2905
QY      285  lnuGluLeuIleLysGluThrLeuGluLeuGluLysCysLeuSerArg----- 300
Db      2906  AGGAAGCTGCCAGCAGAAAGCTGCACACTTGAGAAAGGTCAAGGCTGAGGCCAAGATCAAGA 2965
QY      301  -----MetGluAspGluAsnAsnArgLeuArgLeuGluSerL 313
Db      2966  AACTGAGAGATGATCTCTGTCATGATGATCAGACATATAACTATCAAAAGAACGAA 3025
QY      313  ysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuLeu-----AspA 331
Db      3026  AACTTCCTT-----GAGGAGAGGATTAGTACTTAACGACCAATCTTGCAGAGAGG 3076
QY      331  rGLeuArgAlaGluAsnLeu-----GlnLeuLeuThrGluA 343
Db      3077  AAGAAAGGCCAGAGATCTTACCAAGCTGAAAAACACAGCATGATCTATGATTTCAAGAC 3136
QY      343  snGluLeuHisArgGlnGlnGlu-----ArgAlaProLeuSerLys 356
Db      3137  TGGAACTCGCGCTAAAGAAAGAGAAAGACCGCAGAGAGCTGGAGAG 3186
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_p1us_p2n model

Run on: June 17, 2003, 09:52:48 : Search time 3255 seconds
(without alignments)
3209.803 Million cell updates/sec

Title: US-09-972-758a-2
Perfect score: 1910
Sequence: 1 MAEPFLSEYOHOPQTSNCTG.....LTENELHROQERAPLSKFGD 359

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_p2n.model -BEV=x1h
-Q=/cg2.1/uspto.spool/us09972758/runat.17062003.095237.24371/app.query.fasta.1.519
-DB=GenEmbl -QFMT=fastop -SUFFIX=range -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=us09972758@cgn.1.1.2496 -runat.17062003.095237.24371 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
GenEmbl: *
1: gb_ba: *
2: gb_hg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *

29: em_vl: *
30: em_hg_hum: *
31: em_hg_in: *
32: em_hg_other: *
33: em_hg_mus: *
34: em_hg_pln: *
35: em_hg_rod: *
36: em_hg_jam: *
37: em_hg_vrt: *
38: em_sy: *
39: em_hggo_hum: *
40: em_hggo_mus: *
41: em_hggo_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	2086	9 BC006460	BC006460 Homo sapi
2	1910	100.0	3624	9 AB021179	AB021179 Homo sapi
3	1910	100.0	100836	2 AC087298	AC087298 Homo sapi
4	1903	99.6	2178	9 AK023624	AK023624 Homo sapi
5	1858.5	97.3	142326	2 AC024047	AC024047 Homo sapi
6	1587.5	83.1	3402	10 AY090614	AY090614 Mus muscu
7	1587.5	83.1	111002	10 AL731805	AL731805 Mouse DNA
8	1577.5	82.6	1488	10 BC022111	BC022111 Mus muscu
9	1514.5	79.3	175120	2 AC120950	AC120950 Rattus no
10	629	32.9	461	6 AX331371	AX331371 Sequence
11	556.5	29.1	1330	9 AK056946	AK056946 Homo sapi
12	550	28.8	116035	2 AC008105	AC008105 Homo sapi
13	550	28.8	223442	10 AL662804	AL662804 Mouse DNA
14	545.5	28.6	1238	9 BC025970	BC025970 Homo sapi
15	545	28.5	2048	9 AK096785	AK096785 Homo sapi
16	214	11.2	1536	3 AY051786	AY051786 Drosophill
17	205.5	10.8	183439	3 AC007809	AC007809 Drosophill
18	205.5	10.8	232744	3 AE003705	AE003705 Drosophill
19	203.5	10.8	237119	2 AC017740	AC017740 Drosophill
20	187	9.8	165197	2 AC112538	AC112538 Rattus no
21	183	9.6	59599	2 AC107153	AC107153 Rattus no
22	170	8.9	9121	4 OCTRICHA	Z19092 O. cuniculus
23	165.5	8.7	9344	4 OATRICH	Z18361 O. aries tri
24	165	8.6	6644	4 RABMCP	M77812 Rabbit myos
25	164.5	8.6	4143	5 GSDYNACT	X62773 Gallus gall
26	162	8.5	4152	9 HUMMYH9	M81105 Homo sapien
27	162	8.5	278652	2 AC073809	AC073809 Mus muscu
28	160	8.4	241928	2 AC123859	AC123859 Mus muscu
29	159	8.3	151578	9 AL589986	AL589986 Human DNA
30	159	8.3	256346	2 AC105970	AC105970 Mus muscu
31	158.5	8.3	9551	6 AR076233	AR076233 Sequence
32	158.5	8.3	9551	6 I39845	AC076233 Sequence
33	158.5	8.3	9551	6 HUMRYVAL	I39845 Sequence 93
34	158.5	8.3	126357	2 AC121480	L09190 Human trich
35	157.5	8.2	74317	8 NC12F11	AC121480 Rattus no
36	157	8.2	174241	9 AC018634	AL451017 Neurospor
37	156	8.2	170778	8 OSJN00052	AC018634 Human Chr
38	156	8.2	174019	2 AC027267	AL056627 Oryza sat
39	155	8.2	185155	2 AC073446	AC027267 Homo sapi
40	155.5	8.1	5122	9 HUMMYNM	AC073446 Homo sapi
41	155.5	8.1	174472	2 AC074356	M31013 Human nomu
42	155	8.1	2277	6 AR004983	AC074356 Mus muscu
43	155	8.1	5002	10 DB9728	AR004983 Sequence
44	155	8.1	143487	2 AC016033	DB9728 Mus muscu
45	155	8.1	173934	2 AC100756	AC016033 Homo sapi
					AC100756 Homo sapi

RESULT 1

ALIGNMENTS

BC006460
LOCUS BC006460 2086 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, HMBA-inducible, clone MGC:1880 IMAGE:3535529, mRNA,
complete cds.
ACCESSION BC006460
VERSION BC006460.1 GI:13623668
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2086)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapsb-email.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rudin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susana Chan, Readman Chiu, Chris Fjell, Erin Galland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeay, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Scott,
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 8 Row: d Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4062855.
FEATURES
Source
Location/Qualifiers
1..2086
/organism="Homo sapiens"
/db_xref="LocusID:10614"
/db_xref="taxon:9606"
/clone="MGC:1880 IMAGE:3535529"
/issue_type="Lung, small cell carcinoma"
/clone_1ib="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: POT87"
606..1685
/codon_start=1
/product="HMBA-inducible"
/protein_id="AAH06460.1"
/db_xref="GI:13623669"
/translation="MAPELSEYOHOPOTSNGTGAAYOELNPPKPAEERPEED
SRMOSRAPOLGRRPGEGLSLQPPLOTQACPESSCLRGCEGSDSSAGD
FPPPAVEPTPEELLAPQCHDESKLAPPAAGSEBNGOOROLGKKHRRBPK
KRHKPYKLTWEKKKFKDEKSLRASRIAEKFAQVPAPYNTTQFLMDHDOEP
DLKTGLYSKRAAKSDTSDDDFMEEGSDGMDGDSFLLORDFSERYHT
ESLONMSKQELIKLEYLEKCLSRMEDENRRLRLESKRLGDDARVRELELEIDRLRA
ENLQLTENLHROERAPLSKRGD"
BASE COUNT 510 a 505 c 618 g 453 t
ORIGIN
Alignment Scores:
Pred. No.: 6, 06e-93 Length: 2086
Score: 1910.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-972-758a-2 (1-359) x BC006460 (1-2086)
QY 1 MetAlaGluProPheLeuSerGluTyrGlnHISGlnProGlnThrSerAsnCyThrGly 20
D 606 ATGGCCGAGCATTCTTGTGACAAATATCAACACACACCTCAAACTGTGACACGT 665
QY 21 AlaAlaValGlnGlnGlnLeuAsnProGlnAArgProGlnAlaGlnGlnAArgVal 40
D 666 GCTGCTGCTGTCCAGGAAAGCTGAACCTGAGCGCCGCCAGAGCGGAGAGCGGGT 725
QY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlnGlyArgProGly 60
D 726 CCCGAGAGAGACAGTAGTGAGCAATGAGAGCGTCCCGACATGGTGAGCTCCGGG 785
QY 61 ProGluGlyGlnGlySerLeuGlnGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
D 786 CCGAGGGGGAGAGGAGCGCTGGAATCCCAACCACTCCCTTGACAGCCAGGCTTCCA 845
QY 81 GluSerSerCysLeuArgGlnGlnGlyGlyGlnGlnAsnGlyAspAspSerAlaGly 100
D 846 GAATCTAGCTGCTGTGAGAGAGCGGCGAGAGGCCAATGAGGACACTGCTCCGCTGGC 905
QY 101 GlyAspPheProProProAlaGlnValGluProThrProGluAlaGlnLeuLeuAlaGln 120
D 906 GCGCACTCCCGCCCGCCGAGAAAGTGGAGCCGAGCCGAGGCGAGCTGCTCGCCAG 965
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlnAlaProAlaAlaGlyGlnGlnGlu 140
D 966 CCTTGTGATGATCTCCAGGCGCAATGATGGGGGCTCTCCCGAGGGGGCGAAGAGAG 1025
QY 141 TrpGlnGlnGlnGlnArgGlnLeuGlnGlyLysLysHisArgArgArgProSerLysLys 160
D 1026 TGGGACAGCAGCAGACAGACAGCTGGGGGAAAAAATATGAGAGAGCCCTTCAAGAG 1085
QY 161 LysAlaGlnHisTrpLysProTyrTyrLysLeuThrTrpGlnGlnLysLysLysPheAspGlu 180
D 1086 AACCGGCAATGGAAACCGTACTCAAGCTGAGCTGGAGAGAGAAAAAATGCTCAGCAG 1145
QY 181 LysGlnSerLeuArgAlaSerArgGlnLeuArgAlaGlnPheAlaLysGlnGlnProVal 200
D 1146 AAACAGAGCCTTGACCTTCAAGATCCAGCCGCAATGCTGCCAAGGCGCACGCCGGTC 1205
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnLupProAspLeu 220
D 1206 GCGCCCTTAATACACCAAGCAGGATCTCTCATGATGATGATGACAGACAGGAGCGGACCTC 1265
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240
D 1266 AAACCGCGCTGTACTCCAAAGCGGCGCCGCAATCCGACGACACACGCGATGACGAC 1325
QY 241 PheMetGlnGlnGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
D 1326 TTCTATGAGAGAGAGGGGGTGAAGAGATGGGGGAGCGATGGAGTGGAGGAGCGGCAC 1385
QY 261 GluPheLeuGlnIleArgAspPheSerGlnThrTyrGlnIleArgTyrHisThrGlnSerLeuGln 280
D 1386 GATTTCTCTGACGGGAGCTTCTGAGACGCTACGACCGGTACACAGGAGACCTCGCAG 1445
QY 281 AsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
D 1446 AACATGAGCAGACAGAGAGCTCATCAAGAGAGTACCTGAGCTGAGAGAGAGCTCTCGCCG 1505
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlyLysAspAla 320
D 1506 ATGAGAGAGCAGAGAACCGGCTGCGTGGAGAGCAACCGCTGGTGGCAGCAGCGG 1565
QY 321 ArgValIleArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
D 1566 CCGTCTCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAG 1625
QY 341 ThrGluAsnGlnLeuLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
D 1625

Db 1626 ACCGAGACGAACCTGCACCGGAGAGACGCGCTTTCGAAGTTGGAGAC 1682

RESULT 2
LOCUS AB021179 3624 bp mRNA linear PRI 05-DEC-2000
DEFINITION Homo sapiens mRNA for HEXIM1 protein, complete cds.
ACCESSION AB021179.1 GI:4062855
VERSION AB021179.1
KEYWORDS HEXIM1; HEXIM1 protein.
SOURCE Homo sapiens coronary artery smooth muscle cell cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kusunohara,M., Nagasaki,K., Kimura,K., Maass,N., Manabe,T.,
Ishikawa,S., Aikawa,M., Miyazaki,K. and Yamaguchi,K.
TITLE Cloning of hexamethylene-bis-acetamide-inducible transcript,
HEXIM1, in human vascular smooth muscle cells
JOURNAL Biomed. Res. 20, 273-279 (1999)
REFERENCE 2 (bases 1 to 3624)
AUTHORS Kusunohara,M.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1998) Masatoshi Kusunohara, National Defense
Medical College, The First Department of Internal Medicine; 3-2
Namiki, Tokorozawa, Saitama 359-8513, Japan
(E-mail:mkusun@me.ndmc.ac.jp, Tel:+81-42-995-1597,
Fax:+81-42-996-5200)

FEATURES
source location/Qualifiers
1..3624 /organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="coronary artery smooth muscle cell"
1..3624 /gene="HEXIM1"
690..1769 /gene="HEXIM1"
/codon_start=1
/product="HEXIM1 protein"
/protein_id="BA36166.1"
/db_xref="GI:4062855"
/translation="MAEPFLSEYOHQPOHSNCTGAANOBEINPERPPAGEEVRPED
SRKOSRAFPQLGCRPGEBEGSLESPPPLDTQACPESSCLREGEQMGDDSSAGCD
FPPAEVEPTPEAEELIAPCHDSEASKLAPAGAGEEENQQRQLGKRHRSPSK
KRHWKPYRLTWTEKKRFDKQSLRAKSRIRAEFMAGQVAPYNTQFLMDHDEP
DLKTGLYSRAAKASDSDTDDFMEEGEGDSDGCGDSFLORDESEYERYHT
ESTJONSKOELIKEYLEKLSRMEDENNRLRLFSKRLGGDDARVBELELDRLRA
ENLILTEMLHROERAPLSKRGD"

BASE COUNT 978 a 775 c 964 g 907 t
ORIGIN

Alignment Scores:
Pred. No.: 1..06e-92 Length: 3624
Score: 1910.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-972-758a-2 (1-359) x AB021179 (1-3624)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCystrhngly 20
Db 690 ATGGCCGAGCCATTCTTGTTCAGAAATCAACACACCCCTCAACTAGCACTGTACAGGT 749
QY 21 AlaAlaAlaValGlnGlnGluLeuAsnProGluArgProGlnLysAlaGlnGluArgVal 40
Db 750 GCTGCTGCTGCTCCAGGAAGAGCTGAACCCCTGAGCGGCCCCAGCGGAGGCGCGTG 809
QY 41 ProGlnGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyArgProGly 60
Db 810 CCGGAGAGAGACAGTAGTGCGCAATGAGAGAGCTTCCCGCCAGTTGGTGCGCCGTCCGGG 869
QY 61 ProGlnGlyGlnGlySerLeuGlnSerGlnProProProLeuGlnInThrGlnAlaCysPro 80

Db 870 CCGGAGGGGAAAGCGAGCCTGGAAATCCCAACCACTCCCTTGCAGACCCAGGCTGTCCA 929
QY 81 GluSerSerCysLeuArgGlnGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
Db 930 GAATCTAGCTGCTCTGAGAGAGGGGCGAAGGGGCGACAAATGGGGACACTGTCCTGCTGC 989
QY 101 GlyAspPheProProProAlaGluValGluProThrProGlnAlaGluLeuLeuAlaGln 120
Db 990 GCGGACTTCCCGCGCGCGGCGAGACAGTGAACCGACGCCGAGGCGCAGCTGCTCGCCGAG 1049
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlnGlyGlnGlnGly 140
Db 1050 CCTTGTCATGACTCCGAGGCGCACTAGATTGGGGGCTCTGCGCCAGGCGGCGAAGAGAG 1109
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysHisArgArgArgProSerLysLys 160
Db 1110 TGGGACACACAGCAGAGAGACGCTGGGGGAAAAACATAGGAGAGCGCCGCTCAAGAG 1169
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGlnGlnLysLysLysPheAspGlu 180
Db 1170 AACCGGCATTGGAAACCGTACTACAGCTGACCTGGGAAGAGAGAAAAAGTTCCAGCGAG 1229
QY 181 LysGlnSerLeuArgAlaSerArgLleArgAlaGluMetPheAlaLysGlyGlnProVal 200
Db 1230 AAGACAGAGCCTTCGAGCTTCAAGAGATCCGAGCGAGAGATGTTGCCCAAGGCGCAGCGTC 1289
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnGluProAspLeu 220
Db 1290 GCCCCCTAATACACCCAGCAGAGTTCCCTCATGAGATGATCAGACACGAGAGCGAGCTTC 1349
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240
Db 1350 AAAACCGGCTGTACTCTCAACCGCGCGCGCCCAATTCGACAGCACCCAGCGATGACGAC 1409
QY 241 PheMetGlnGlnGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
Db 1410 TTCATGGAAGAAAGGGGCGTAGAGAGATGGGGGAGCAGCATGGATGGAGAGGAGCGAC 1469
QY 261 GluPheLeuGlnArgAspPheSerGluTyrTrpGluArgTyrHisThrLysLeuGln 280
Db 1470 GAATTTCTCAGGGGAGCTTCTCGAGACGTACGAGCGGTACACAGGAGAGCCCTGCAG 1529
QY 281 AsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
Db 1530 AACATGACGACGAGGAGAGCTCATCAGCACTACTGTAAGTGAAGAGGCTCTGCGCC 1589
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlyLysAspAla 320
Db 1590 ATGGAGAGCAGAGAACACCGGCTGCGGTGAGAGCAACGCGCTGCGGCAGCACCGCG 1649
QY 321 ArgValArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
Db 1650 CGTGTCCGGAGCTGGAGCTGGAGCTGGACCGGCTCGCGCGAGAACTCCAGCTGCTG 1709
QY 341 ThrGlnAsnGlnLeuLeuLysArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
Db 1710 ACCGAGACGAACCTGCACCGCGCAGCAGAGCGCGCGCTTTCGAAGTTTGACAGAC 1766

RESULT 3
AC087298 100836 bp DNA linear HTG 20-APR-2002
LOCUS AC087298 Homo sapiens chromosome 17 clone RP11-40363 map 17, WORKING DRAFT
DEFINITION SEQUENCE: 15 ordered pieces.
ACCESSION AC087298 GI:20219131
VERSION AC087298.8
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 100836)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 17, clone RP11-403G3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 100836)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gadyana, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Lahocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 100836)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gadyana, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamal, A., Karatas, A., Kells, C., Lahocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 20, 2002 this sequence version replaced gi:20177719. All repeats were identified using RepeatMasker:
Shit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: LI2020

Center clone name: 403.G.3

Summary Statistics

Sequencing vector: Plasmid; n/a: 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 96277 bases at least Q40

Consensus quality: 98051 bases at least Q30

Consensus quality: 98838 bases at least Q20

Insert size: 14000; agarose-fp

Insert size: 99436; sum-of-contigs
Quality coverage: 19.1 in Q20 bases; agarose-fp
Quality coverage: 26.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter.

This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.

1 8098: contig of 8098 bp in length

8099 8198: gap of 100 bp

8199 8271: contig of 73 bp in length

8272 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

8372 8371: gap of 100 bp

FEATURES

Source

1. 100836

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="17"

/map="17"

/clone="RP11-403G3"

/clone_lib="RP11 Human Male BAC"

1. 8098

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

8199. 8271

/note="assembly_fragment"

8372. 9497

/note="assembly_fragment"

9598. 10863

/note="assembly_fragment"

10964. 12232

/note="assembly_fragment"

12333. 13844

/note="assembly_fragment"

13945. 15215

/note="assembly_fragment"

15316. 18234

/note="assembly_fragment"

18335. 21787

/note="assembly_fragment"

21888. 25485

/note="assembly_fragment"

25586. 40531

```
misc_feature /note="assembly_fragment"
40632..52119
/note="assembly_fragment"
misc_feature 52220..64841
/note="assembly_fragment"
misc_feature 65042..79587
/note="assembly_fragment"
misc_feature 79688..100836
/note="assembly_fragment"

BASE COUNT 23052 a 27082 c 27053 g 22249 t 1400 others
ORIGIN

Alignment Scores:
Pred. No.: 3e-91 Length: 100836
Score: 1910.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 2

US-09-972-758A-2 (1-359) x AC087298 (1-100836)

QY 1 MetalagluPProPheLeuSerGluTyrGlnHisGlnProGlnInThrSerAsnCysThrGly 20
Db 70785 ATGGCCGAGCCATTCTTGTGCAAAATATCAACACCGCCTCAACACTGACGTACAGGT 70726
QY 21 AlaAlaIaValGlnGluGluLeuAsnProGluArqProProGlyAlaGluGluVal 40
Db 70725 GCTGCTGCTGCCAGAAAGACGTGACCTGAGCCGCCCCAGCGCGGAGAGCGGGTG 70666
QY 41 ProGluGluAspSerArqTrpGlnSerArqAlaPheProGlnLeuGlyIaArpProGly 60
Db 70665 CCGGAGAGAGACAGTACGTGCGCAATCGAGACGCTTCCCGAGTTGGTGCGCTCCGGG 70606
QY 61 ProGluGlyGluGlySerLeuGluSerLInProProProLeuGlnInThrGlnAlaAspPro 80
Db 70605 CCGGAGGGGGAAGGAGGCTTGAAATCCACACCACTCTCTTGACACACCGCCTGTCCA 70546
QY 81 GluSerSerCysLeuArqGluGluGluGlyGlnAsnGlyAspAspSerSerIaGly 100
Db 70545 GAATCTACTGCTGCTGAGAGAGGGCGGAGGCGCGAATGGGAGACTCTGCTCCCTGCG 70486
QY 101 GlysPhePProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
Db 70485 GGGGACTTCCCGCCCGCGAGAAAGTGAACCGAGCGCCGAGCGGCGACTCTGCCAG 70426
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaIaGlyGluGluGlu 140
Db 70425 CTTGTCTATGACTCCGAGGCCAGTAAATTGGGGGCTCTCTGCCCGGAGGCGCAAGAGAG 70366
QY 141 TrpGlyGlnGlnIaArqGlnLeuGlyLysLysHisArqArqArqProSerLysLys 160
Db 70365 TGGGAGACGACGACAGAGACAGTGGGGAAGAAAAACATAGAGAGCGCCCTCCAGAGAG 70306
QY 161 LysArqHisTrpLysProTyrTrpLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
Db 70305 AAGCGGCAATTGCAAAACCTACTACAACTGACCTGCAAGAGAAAGAAAGTTCAGCAGAG 70246
QY 181 LysGlnSerLeuArqAlaSerArqGlnLeuArqAlaGluMetPheAlaLysGlyGlnProVal 200
Db 70245 AAACGAGACCTTCGAGCTTCAAGATCCGAGCCGAGATGTTCCCAAGGGCCAGCGGCTC 70186
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
Db 70185 GCGCCCTTAAACACGACGACGCTTCTCATGATGATCATCAGACGAGAGAGACCGGATCTC 70126
QY 221 LysThrGlyLeuTyrSerLysArqAlaAlaLysSerAspAspThrSerAspAspAsp 240
Db 70125 AAAACCGGCTGTACTCCACAGCGGCGCCGCCCAATATCCAGACACGATGATGACAGC 70066
QY 241 PheMetGluGluGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 260
Db 70065 TTCAATGAGAAAGGGGGTGGAGGATGGGGCACCGATGGGATGGAGGGAGCGACGCGACG 70006
```

```
QY 261 GluPheLeuGlnArqAspPheSerGluThrTyrGluArqTyrHisThrGluSerLeuGln 280
Db 70005 GAGTTTCTGCGAGCGGACTTTCGGAGACGATACAGCGGTATCCACACGAGAGGCTTCGAC 69946
QY 281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluGluGluGluGluGluGluGlu 300
Db 69945 AACATGAGCAACGAGAGGCTCATCAGAGATCTGGAACATGGAGAGATGGCTCTCCGCC 69886
QY 301 MetGluAspGluAsnArqArqLeuArqLeuGluSerLysArqGluGlyIaAspAla 320
Db 69885 ATGAGAGAGCAACACCGGCTGCTGGAGAGCAAGCGGTGGGGGAGGAGCGCG 69826
QY 321 ArqValArqGluLeuGluGluLeuGluAspArqLeuArqAlaGluAsnLeuGluLeu 340
Db 69825 CGTGTGCGGAGCTGAGACTGAGACTGAGACCGGCTGCCCGCGAGAACCTGACACTGCTG 69766
QY 341 ThrGluAsnGluLeuHisArqGlnGlnGluArqAlaProLeuSerLysPheGlyAsp 359
Db 69765 ACCGAGAAAGAACTGACCGGCGAGAGAGCGCGGCTTCCAAAGTTTGAGAGAC 69709

RESULT 4
AK023624 2178 bp mRNA linear PRI 01-AUG-2002
LOCUS Homo sapiens cDNA FLJ13562 fis, clone PLACE1008080, highly similar
DEFINITION to Homo sapiens mRNA for HEXIM1 protein.
ACCESSION AK023624
VERSION AK023624.1 GI:10435606
KEYWORDS Oligo capping; fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_l1b:PLACE1
clone:PLACE1008080.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Negai,K., Sugano,S., Shitatori,A., Sudo,H.,
Wagatsuma,M., Hosofri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahata,K., Masuo,Y., Niomiya,K. and Iwayanagi,T.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 2178)
JOURNAL Isogai,T. and Otsuki,T.
REFERENCE Direct Submission
AUTHORS Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Source location/Qualifiers
1..2178
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1008080"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="Cloning vector: PME18SFL3"

BASE COUNT 525 a 513 c 678 g 462 t
ORIGIN

Alignment Scores:
Pred. No.: 1.49e-92 Length: 2178
Score: 1903.00 Matches: 358
Percent Similarity: 99.72% Conservative: 0
Best Local Similarity: 99.72% Mismatches: 1
```

Query Match: 99.63% Indels: 0
 DB: 9 Gaps: 0
 us-09-972-758a-2 (1-359) x AK023624 (1-2178)

QY 1 MetAlaGluProPheUeuSerGluYrGlnHisGlnProGlnThrSerAnCySthrcly 20
 DB 716 ATGGCCGACCATTTCTTGTGAGATATATCAACACCCCTCAAACTAGCACTGTACAGGT 775
 QY 21 AAlaAlaValaGlnGluGluLeuAnProGluArProGluYAlaGluGluArGVal 40
 DB 776 GGTGCTGTCTCAGAGAGAGTGAACCTGAGCCGCCCAAGCGCGAGAGAGCGGGT 835
 QY 41 ProGluGluAspSerArGTrpGlnSerArGAlaPheProGlnLeuGlyLarProGly 60
 DB 836 CCGGAGAGGACAGTAGTGCGCAATCGAGAGCGTTCCCTCCAGTTGGTGGCCGCGGG 895
 QY 61 ProGluGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnIlaCyPro 80
 DB 896 CCGGAGGGGAAAGGAGCTGGAAATCCCAACCACTCCCTCCAGACCAGCCCTGTCCA 955
 QY 81 GluSerSerCysLeuArGluGlyLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
 DB 956 GAATCTAGCTGCTGAGAGAGGCGGAGAGGCGCAAGATGGGAGAGACTGCTCCCTGCG 1015
 QY 101 GlyAspPheProProProAlaGluValaGluProThrProGluAlaGluLeuAlaGln 120
 DB 1016 GCGGACTTCCCGCCCGGACAGATGTGAACCGACCCCGAGCCGAGCTGCTCCCGCAG 1075
 QY 121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGluGlu 140
 DB 1076 CCTTGCTGACTCCGAGAGCCGTAAGTTGGGGGCTCCTCGCCGAGGGGCGCAAGAGGAG 1135
 QY 141 TrpGlyGlnGlnGlnArGlnLeuGlyLysLysHisArGArGArProSerLysLys 160
 DB 1136 TGGGACAGCAGACAGACAGCTGGGGAAGAAAAACATAGAGAGCGCCGTCGAAGAAG 1195
 QY 161 LysArGHisTrpLysProGlyTrpLysLeuThrTrpGluGluLysLysLysLysGlu 180
 DB 1136 AAGCGGCAATGGAAACCGTACTACAACTGACCTGGGAAGAAAGAAAGTTTCACAGAG 1255
 QY 181 LysGlnSerLeuArGAlaSerArGAlaGluMetPheAlaLysGlyGlnProAla 200
 DB 1256 AAGCAGAGGCTTCGAGCTTCAGAGATCCGAGCCGAGATGTTGCCAAGGCGCAAGCGGTC 1315
 QY 201 AlaProTyraSerThrThrGlnPheLeuMetAspHisAspGlnGluProAspLeu 220
 DB 1316 GCGCCCTATACACACCAGCTTCTCATGTGATGATCAAGACAGAGAGCGGATCTC 1375
 QY 221 LysThrGlyLeuTySerLysArGAlaAlaLysSerAspPthSerAspAsp 240
 DB 1376 AAGACCGGCTGTACTCCAAAGCGGGCGCCCGCAATCCGAGCAGACAGCATACAGAC 1435
 QY 241 PheMetGluGluGlyGlyGluGluAspGlyLysSerAspGlyMetGlyGlyAspGlySer 260
 DB 1436 TTCATGGAAGAAAGGGGAGAGATGGGGCAGCCATGGGATGGAGGGAGCGCGAGC 1495
 QY 261 GluPheLeuGlnArGAspPheSerGluThrTyrgluArGlyHisThrGluSerLeuGln 280
 DB 1496 GAGTTTCTGACAGCGGAGCTTCTCGGAGAGTACGAGCGGTACACACAGAGAGCTGAG 1555
 QY 281 AsnMetSerLysGlnGluLeuLeuLysGlnLysLysLysLysLysLysLysLysLys 300
 DB 1556 AACATGAGCAAGCAGAGCTATCAAGAGTACCTGGAACCTGGAAGAACTGCTCCGCC 1615
 QY 301 MetGluAspGluAsnAsnArGLeuArGLeuGluSerLysArGLeuGlyGlyAspAspAla 320
 DB 1616 ATGGAGGAGCAAGCAACCGGCTGGGCTGGAGAGCAAGCGGCTGGGCTGGAGAGCCG 1675
 QY 321 ArGValArGluLeuGluGluLeuAspArGLeuArGAlaGluAsnLeuGlnLeuLeu 340
 DB 1676 CGTGTGCGGAGAGCTGAGACTGAGACTGAGACCGGCTGCGCGGAGAACTCTCAGCTGTG 1735

QY 341 ThrGluAsnGluLeuHisArGlnGlnGluArGAlaProLeuSerLysPheGlyAsp 359
 DB 1736 ACCGAGAACGACTGACCGGAGCAGAGGCGCGCCCTTCCAAAGTTTGAGAC 1792

RESULT 5
 AC024047/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

On Jun 16, 2000 this sequence version replaced gi:7109599.

Center: Washington University Genome Sequencing Center
 Web site: http://genome.wustl.edu/gsc/index.shtml

Project Information
 Center project name: H.NH0403G03

Summary Statistics

Sequencing vector: MJ3; 100%
 Sequencing method: Plasmid; 0%
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 127642 bases at least Q40
 Consensus quality: 133308 bases at least Q20
 Consensus quality: 135734 bases at least Q20
 Insert size: 147000; agarose-fp
 Insert size: 139626; sum-of-ctrls
 Quality coverage: 4.06 in Q20 bases; agarose-fp
 Quality coverage: 4.33 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1575: contig of 1575 bp in length
 1576 1675: gap of unknown length
 1676 3114: contig of 1439 bp in length
 3115 3214: gap of unknown length
 3215 5389: contig of 2175 bp in length
 5390 5489: gap of unknown length
 5490 7129: contig of 1640 bp in length
 7130 7229: gap of unknown length
 7230 9240: contig of 2011 bp in length
 9241 9340: gap of unknown length
 9341 11177: contig of 1837 bp in length
 11178 11277: gap of unknown length
 11278 13113: contig of 1836 bp in length
 13114 13213: gap of unknown length
 13214 16526: contig of 3313 bp in length
 16527 16626: gap of unknown length
 16627 19864: contig of 3238 bp in length

PUBMED 12119119
2 (bases 1 to 3402)
REFERENCE
AUTHORS Huang, F., Wagner, M. and Siddiqui, M.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2002) Anatomy and Cell Biology, State University
of New York Health Science Center, 450 Clarkson Avenue, Brooklyn,
NY 11203, USA

FEATURES

source

gene
CDS

Location/Qualifiers
1..3402
/organism="Mus musculus"
/strain="129/SvJ"
/db_xref="taxon:10090"
1..3402
/gene="Clp1"
596..1666
/gene="Clp1"
/note="nuclear protein; CLP-1"
/codon_start=1
/product="cardiac lineage protein 1"
/protein_id="AA09026.1"
/db_xref="GI:20135641"
/translation="MAEPLLEHGHQHPQTSNCTGAAYVHEHTSERPPSAEERPKED
SRMQSRALQSGSPGEGGLKHPLEPLTNACPELSLEKGGKONGDLSTGA
SPSAEGEMSESLVOPGHDSPATKQEAAPAGGEPWGOOROLGKKRRRPSKKRH
WKPYKLTMEKKKKFDEKQSLASRVRAEMAKGQPVAPYNTOTFMDHQBEPDLK
TGLYPRAAKSDTSDDEYVEKEGEDGSDGMDSEFLQRPSEYERYHAESL
QMSKQELIKELKLSRKEDENNRLRLEKRLGVDARVRELELELDRLRAENT
QLTLENELHROERAPLSKFGD"
BASE COUNT 920 a 771 c 890 g 821 t
ORIGIN

Alignment Scores:

Pred. No.: 1.19e-75 Length: 3402
Score: 1587.50 Matches: 308
Percent Similarity: 88.86% Conservative: 11
Best Local Similarity: 85.79% Mismatches: 37
Query Match: 83.12% Indels: 3
DB: 10 Gaps: 2

US-09-972-758a-2 (1-359) x AY090614 (1-3402)

QY 1 MetAlaGluPProPheLeuSerGluTyrGlnHISGlnProGlnThrSerAsnGlyThrGly 20
DB 596 ATGCCCGAGCCACTTGTACAGAACATCAACACCCGCTCAACATGACACTGTACAGT 655
QY 21 AAlaAlaValAlGlnGluGluLeuAsnProGluArgProProGlyAlaGluGluVal 40
DB 656 GCTGCTGTTCATGAGAGACATACCTCTGAGCGCCCCCAAGCGCGAGAGAACGGGTG 715
QY 41 ProGluGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
DB 716 CCCAAGGAGACAGTACGTGGCAACGAGACCGTCTTGCAGTCCGCTAGCCGTCCACGG 775
QY 61 ProGluGluGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
DB 776 CAGAGAGGAGAGAGGGCGCTGACAGACCACTGCCCATTCGACAGCAATGCTGTCCA 835
QY 81 GluSerSerCysLeuArgGluGlyGlyLysGlnAsnGlnGlyAspAspSerSerAlaGly 100
DB 836 GAATTGAGCTCCCTGAGAGAGGGCGAGAAAGGCCAGATGGGAGAGACTTATCCACTGTC 895
QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
DB 896 GGT---GCCTCCCGTCCGCGAGAGAGACCGCATGTGACAGATCC-----CTCGTGCAG 946
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGluGluGlu 140
DB 947 CCAGGTCTACTGACTGGAGGCCACCAAGAGAGAGGCTCTGCGCGGAGGCGAGAGCCA 1006
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
DB 1007 TGGGAGACGCAACAGAGACACCTGGCGCAAGAAAAACATCGAGAGACGCCCTCAAGAGAG 1066

QY 161 LysArgHisTrpLysProTyrTrpLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
DB 1067 AAGCGCATTTGGAGACCCCTACTACAGCTGCTGGAGGAGAGAAAAAGTTGACAGCAG 1126
QY 181 LysGlnSerLeuAlaGlnAlaSerArgLysArgAlaGluMetPheAlaLysGlyGlnProVal 200
DB 1127 AAGCAGACCTTCGCGAGCTTCCGGGTTGACCCCGAGATGTTGCGCAAGGGCCACCACTT 1186
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetLysAspHisAspGlnGluGlnProAspLeu 220
DB 1187 GCGCCCTATACACACACCACTTCTCTGATGATGACCAACGATCAGAGAGACCTGATCTC 1246
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240
DB 1247 AAAACCGGCTTTATCCCAACCGGCGACCCCAATATCGACACACAGCATGAGAGAT 1306
QY 241 PheMetGluGluGlyGlyGluGluAspGlyLysSerAspGlyMetGlyLysGlySer 260
DB 1307 TTTGTGAGAGAACTGTGTGAGAGAGACGAGGACCGATGATGATGAGAGGAGGAGCGCAGC 1366
QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
DB 1367 GAGTTTCTGACGCGGACTTCTCGGAGACGTACGAGCGGCTACCAACGCGAGAGCTGAG 1426
QY 281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluGluLysCysLeuSerArg 300
DB 1427 AACATGACCAAGCAGGACCTCATCAACAGTACCTGGAGCTGGAGAGAGTGCCTCCCGC 1486
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyLysAspAspAla 320
DB 1487 AAGCAGAGCAAAATACCGGCTGCGGTGAAAGCAAGCGGCTGGAGGCGTGTGAGCGG 1546
QY 321 ArgValArgGluLeuGluGluLeuGluAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
DB 1547 CGAGTTCGCGAGCTGACACTGAGCTGAGCGGCTGCGGCTGAGAACCTCAGCTGCTG 1606
QY 341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
DB 1607 ACCGAGAACGAACGTGACCGCGACAGAGAGCGCGCTTCTTCCAGATTGCGCGCAC 1663

RESULT 7
AL731805
LOCUS Mouse DNA sequence from clone Rp23-358E19 on chromosome 11,
DEFINITION complete sequence.
ACCESSION AL731805
VERSION AL731805.8 GI:21540125
KEYWORDS HMG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Dunn, M.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humgurey@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 23, 2002 this sequence version replaced gi:21531483.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humgurey@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate

chemistry covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP. Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep. RP23-358E19 is from the RCG1-23 Mouse PAC library constructed by the group of Plier de Jong. For further details see <http://www.chori.org/Bacpac/home.htm>.
 VECTOR: pBACe3.6.

FEATURES	Location/Qualifiers
source	1. .111002

BASE COUNT	28051	a	27579	c	27135	g	28237	t
ORIGIN								

Alignment Scores:	
Pred. No.:	3,95e-74
Score:	1587.50
Percent Similarity:	88.86%
Best Local Similarity:	85.79%
Query Match:	83.12%
DB:	10
Length:	111002
Matches:	308
Conservative:	11
Mismatches:	37
Indels:	3
Gaps:	2

US-09-972-758A-2 (1-359) x AL731805 (1-111002)

[illegible]

QY	201	AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnIleProAspLeu	220
Db	92031	GGCCCCCTATACACCAAGCATTTCCATGAGAGACACAGATCAGAGAGGCTTATCTC	92090
QY	221	LysThrGlyLeuLysSerLysArgAlaIleAlaLysSerAspAspThrSerAspAsp	240
Db	92091	AAACCGGCGCTTACCCCAAGCGGGCAGCCGCATTCGACGACACCAAGCATGAGAT	92150
QY	241	PheMetGlnGlnGlyLysGlnGlnAspGlnGlyLysSerAspGlyMetGlyLysAspLysSer	260
Db	92151	TTTGTGGAAAGAACCTGTGAGGAGAAAGCAGGACAGCAAGCATGGAGGAGGACCGCAGC	92210
QY	261	GluPheLeuGlnIleArgAspPheSerGluThrTyrGlnAlaTyrTyrHisThrGlnSerLeuGln	280
Db	92211	GAGTTTCTGACGGGAGCTTCTCGGAGACCTACGAGCGGTACCACCGCAGAGCCTCGAG	92270
QY	281	AsnMetSerLysGlnGlnLeuLeuIleLysGluTyrLeuGlnLeuGlnLysCysLeuSerArg	300
Db	92271	AACATGAGCAACGACGAGAGCTCATCAAGAGTACCTGAGACTGGAGAACGCTCTCCGC	92330
QY	301	MetGluAspGlnAsnAspAsnArgLeuAlaGlnGlnIleSerLysArgLeuGlnGlyLysAspAla	320
Db	92331	AACGAGAGACCAAAATTAACGGGCTGGCGGTGGAAAGCAACGGCTGGAGGCGCTGCACCG	92390
QY	321	ArgValAlaGlnLeuGlnIleLeuGlnLeuAspArgLeuArgAlaGlnAsnLeuGlnLeuLeu	340
Db	92391	CGAGTGGGGAGCTCTGAGGTACAGAGCTGGACCGGCTCGCGGTGAGAACCTCCAGCTCGTG	92450
QY	341	ThrGlnAsnGlnLeuLeuHisLysArgGlnGlnGlnIleAlaArgAlaProLeuSerLysPheGlnLysP	359
Db	92451	ACCGAAGACCAACTTGCACCGGAGCGAGGACGAGCGCTCTTTCCAAAGTTCCGGCAGC	92507

RESULT	8
LOCUS	BC022111
DEFINITION	BC022111 1488 bp mRNA linear ROD-07-AUG-2002 Mus musculus, similar to HmBa-inducible, clone MGC:36249.
ACCESSION	BC022111
VERSION	BC022111.1 GI:18381039
KEYWORDS	MGC.
SOURCE	house mouse. <i>Mus musculus</i> .
ORGANISM	Eukaryote; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1488) Straussberg, R. Direct Submission Submitted (24-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 57 Row: a Column: 15

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency OK analysis, similarity but not identity to protein.

FEATURES

Location/Qualifiers

1..1488

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="CZCH.11"

/clone="MGC:36249 IMAGE:5067620"

/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."

/clone_1ib="NCI CGAP_Lu29"

/lab_host="DH10B"

/note="vector: pCMV-SPORT6"

CDS

/codon_start=1

/product="Similar to HMB-Inducible"

/protein_id="AAH2211.1"

/db_xref="GI:18381040"

/translation="MSESLVDPHDSSEATKQEPAPAGEEPEWQOQROLGKKHRRP
SKKHHMPYKLTWEKKKDEKQSLASRVRAEMFAGQVAPVNTTQFIMDDHQ
EEDPKTGLYPRRAAKSDDTDEDFVEAGEDESDGMGQSEFLOPDFSEYER
YHBSLQWMSKQELTKEYLELEKCSRKEDENNRLRLBSKRLGVDARVRELELDR
LRAENIOLITENEHLHROOEAPVLSKRGD"

BASE COUNT 385 a 371 c 449 g 283 t

ORIGIN

Alignment Scores:

Pred. No.: 1.75e-75 Length: 1488
Score: 1577.50 Matches: 305
Percent Similarity: 88.86% Conservative: 34
Best Local Similarity: 84.96% Mismatches: 37
Query Match: 82.59% Indels: 3
DB: 10 Gaps: 2

US-09-972-758A-2 (1-359) x BC022111 (1-1488)

OY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
DB GTCTCAGCCACTCTTGTGACAGAACATCACACCGCTCAAACTAGCACTGTACAGCT 177
OY 21 AlaAlaValaGlnGluLeuAsnProGluArgProGlyAlaGluGluVala 40
DB GCTGCTGTTTCATGAGACATACCTCTGACGCCGCCCAAGCGCGAGAGAGGGTG 237
OY 41 ProGluGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
DB CCCAAGGAGGACAGTAGGTGCAATCGAGAGCGTCTTCGAGCGCGTAGCGTCCAGG 297
OY 61 ProGluGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
DB CAGAGAGGAGAGGGCGGCTGTAACACACAGCTGCCCTTCGACAGCAATGCTGTCCA 357
OY 81 GluSerSerCysLeuArgGluGluGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
DB GAATTAGGCTCTCGGAGAGAGCGGAGAGAGCGGAGAGAGCTTATTCACGTGGC 417
OY 101 GlyAspPheProProProAlaGluValaGlnProThrProGluAlaGluLeuAlaGln 120
DB GGT---GCCCTCCCGCTGCGGAGAGAGCGGAGAGCGATGTACAGATCC-----CTCGTGCAG 468
OY 121 ProCysHisAspSerGluAlaSerLeuGlnGlyAlaProAlaAlaGlyGluGluGlu 140
DB CCAAGTATGACTCTGAGAGCGGAGAGAGCGGAGAGCGGAGAGAGCGGAGAGAGCA 528
OY 141 TrpGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
DB TGGGACACGACAGACAGACAGCTGGCGCAAGAAAAACATCGAGACGCCCTCAAGAG 588
OY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluGlnGlnGlnGlnGln 180
DB AAGCGCATTGGAAGCCCTACTACAGCTGAGGAGAGGAGAGAGAGAGAGAGAGAGAG 648

OY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
DB AAGCAGAGCCCTGCGAGCTTGGCGGTTGAGCGCGAGATTTGCCAAGGCGCAGCAGTT 708
OY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220
DB GCGCCCTATTAACACAGCAGCTTCTCATGATGACACAGATCAGAGAGAGCTTATCTC 768
OY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240
DB AAAACCGGCTTTTACCAGGCGGAGCGGCGGCAATCCGACACACAGCATAGGAT 828
OY 241 PheMetGluGluGluGlyGluGluGlnAspGlyLysSerAspGlyMetGlyLysPylSer 260
DB TTTGTGAAGAAGCTGTGTAGACAGACGAGCAGCAGATGCCATGGAGGCGGAGCAGC 888
OY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluAlaGlyTyrHisThrGluSerLeuGln 280
DB GAGTTTCTGACCGGAGCTTCTTCGAGACGTAACGAGCGGTACACCGCAGAGCTTCAG 948
OY 281 AsnMetSerLysGlnGluLeuValLeuLysGluTyrGluGluGluGlnGlnGlnGlnGln 300
DB AACATGACGACAGCAGAGCTCTTCAAGAGATCTGAGAGCTGAGAGAGTGCCTTCCCGC 1008
OY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyLysAspAla 320
DB AAGGAGAGCAGAAATTAACCGGCTGCGTGGAACACAGCGGCTGGAGCGCTGCACCG 1068
OY 321 ArgValaArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 340
DB CCAGTGGCGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1128
OY 341 ThrGluAsnGlnGluLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
DB ACCGAGAGACGACATGACCGGCGAGGAGGAGCGGCTTCTTCCAAAGTTGCGGAGC 1185

RESULT 9

AC120950

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 175120)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunay,C., Burch,P., Burrell,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louieged,H.,

REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL
Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Manshwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., McWhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Mosier, M., Neal, D., Newton, J. P., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nookankw, S., Ogih, M., Okunnu, G., Ordunye, N., Overton, R., Pace, A., Payton, B., Peery, J., Perez, L., Peterson, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojudo, A., Rolfe, M., Ruiz, S., Savery, G., Schier, S., Scott, G., Shen, H., Shooshbar, N., Sisson, T., Sodergren, E., Sotak, T., Sparks, A., Stanley, H., Stone, H., Tanton, A., Staley, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telirod, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.	Submitted Unpublished 2 (bases 1 to 175120)	Submitted Unpublished 2 (bases 1 to 175120)	Submitted Unpublished 2 (bases 1 to 175120)
Worley, K. C.	Worley, K. C.	Worley, K. C.	Worley, K. C.
Direct Submission Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 175120)	Direct Submission Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 175120)	Direct Submission Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 175120)	Direct Submission Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 175120)
Worley, K. C.	Worley, K. C.	Worley, K. C.	Worley, K. C.
Direct Submission Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 18, 2002 this sequence version replaced gi:20564294.	Direct Submission Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 18, 2002 this sequence version replaced gi:20564294.	Direct Submission Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 18, 2002 this sequence version replaced gi:20564294.	Direct Submission Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 18, 2002 this sequence version replaced gi:20564294.
----- Genome Center -----	----- Genome Center -----	----- Genome Center -----	----- Genome Center -----
Center: Baylor College of Medicine	Center: Baylor College of Medicine	Center: Baylor College of Medicine	Center: Baylor College of Medicine
Center code: BCM	Center code: BCM	Center code: BCM	Center code: BCM
Web Site: http://www.hgsc.bcm.tmc.edu/	Web Site: http://www.hgsc.bcm.tmc.edu/	Web Site: http://www.hgsc.bcm.tmc.edu/	Web Site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu	Contact: hgsc-help@bcm.tmc.edu	Contact: hgsc-help@bcm.tmc.edu	Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----	----- Project Information -----	----- Project Information -----	----- Project Information -----
Center project name: GWRO	Center project name: GWRO	Center project name: GWRO	Center project name: GWRO
Center clone name: CH230-456H21	Center clone name: CH230-456H21	Center clone name: CH230-456H21	Center clone name: CH230-456H21
----- Summary Statistics -----	----- Summary Statistics -----	----- Summary Statistics -----	----- Summary Statistics -----
Sequencing vector: Plasmid	Sequencing vector: Plasmid	Sequencing vector: Plasmid	Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads	Chemistry: Dye-terminator Big Dye; 100% of reads	Chemistry: Dye-terminator Big Dye; 100% of reads	Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329	Assembly program: Phrap; version 0.990329	Assembly program: Phrap; version 0.990329	Assembly program: Phrap; version 0.990329
Consensus quality: 141564 bases at least Q40	Consensus quality: 141564 bases at least Q40	Consensus quality: 141564 bases at least Q40	Consensus quality: 141564 bases at least Q40
Consensus quality: 147590 bases at least Q30	Consensus quality: 147590 bases at least Q30	Consensus quality: 147590 bases at least Q30	Consensus quality: 147590 bases at least Q30
Consensus quality: 152266 bases at least Q20	Consensus quality: 152266 bases at least Q20	Consensus quality: 152266 bases at least Q20	Consensus quality: 152266 bases at least Q20
----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) * NOTE: This is a 'working draft' sequence. It currently * consists of 41 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.	----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) * NOTE: This is a 'working draft' sequence. It currently * consists of 41 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.	----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) * NOTE: This is a 'working draft' sequence. It currently * consists of 41 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.	----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) * NOTE: This is a 'working draft' sequence. It currently * consists of 41 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.
* 1 1114: contig of 1114 bp in length	* 1 1114: contig of 1114 bp in length	* 1 1114: contig of 1114 bp in length	* 1 1114: contig of 1114 bp in length
* 1115 1214: gap of unknown length	* 1115 1214: gap of unknown length	* 1115 1214: gap of unknown length	* 1115 1214: gap of unknown length
* 1215 2688: contig of 1474 bp in length	* 1215 2688: contig of 1474 bp in length	* 1215 2688: contig of 1474 bp in length	* 1215 2688: contig of 1474 bp in length
* 2689 2788: gap of unknown length	* 2689 2788: gap of unknown length	* 2689 2788: gap of unknown length	* 2689 2788: gap of unknown length
* 2789 4582: contig of 1794 bp in length	* 2789 4582: contig of 1794 bp in length	* 2789 4582: contig of 1794 bp in length	* 2789 4582: contig of 1794 bp in length
* 4583 4683: gap of unknown length	* 4583 4683: gap of unknown length	* 4583 4683: gap of unknown length	* 4583 4683: gap of unknown length
* 4684 6524: contig of 1842 bp in length	* 4684 6524: contig of 1842 bp in length	* 4684 6524: contig of 1842 bp in length	* 4684 6524: contig of 1842 bp in length
* 6525 6625: gap of unknown length	* 6525 6625: gap of unknown length	* 6525 6625: gap of unknown length	* 6525 6625: gap of unknown length
* 6626 8342: contig of 1718 bp in length	* 6626 83		

	*	13600	14781:	contig of 1182 bp in length
	*	14782	14881:	gap of unknown length
	*	14882	17066:	contig of 2185 bp in length
	*	17067	17166:	gap of unknown length
	*	17167	19189:	contig of 2023 bp in length
	*	19190	19289:	gap of unknown length
	*	19290	20786:	contig of 1497 bp in length
	*	20787	20886:	gap of unknown length
	*	20887	22783:	contig of 1897 bp in length
	*	22784	22883:	gap of unknown length
	*	22884	24379:	contig of 1496 bp in length
	*	24380	24479:	gap of unknown length
	*	24480	25904:	contig of 1425 bp in length
	*	25905	26004:	gap of unknown length
	*	26005	28572:	contig of 2568 bp in length
	*	28573	28672:	gap of unknown length
	*	28673	31006:	contig of 2334 bp in length
	*	31007	31106:	gap of unknown length
	*	31107	33737:	contig of 2631 bp in length
	*	33738	33837:	gap of unknown length
	*	33838	35985:	contig of 2148 bp in length
	*	35986	36085:	gap of unknown length
	*	36086	39468:	contig of 3383 bp in length
	*	39469	39569:	gap of unknown length
	*	39569	43008:	contig of 3440 bp in length
	*	43009	43108:	gap of unknown length
	*	43109	45505:	contig of 2397 bp in length
	*	45506	45605:	gap of unknown length
	*	45606	48876:	contig of 3271 bp in length
	*	48877	48976:	gap of unknown length
	*	48977	52243:	contig of 3267 bp in length
	*	52244	52343:	gap of unknown length
	*	52344	56160:	contig of 3817 bp in length
	*	56161	56260:	gap of unknown length
	*	56261	60475:	contig of 4215 bp in length
	*	60476	60575:	gap of unknown length
	*	60576	64722:	contig of 4147 bp in length
	*	64723	64822:	gap of unknown length
	*	64823	68616:	contig of 3794 bp in length
	*	68617	68716:	gap of unknown length
	*	68717	73739:	contig of 5023 bp in length
	*	73740	73839:	gap of unknown length
	*	73840	75346:	contig of 5507 bp in length
	*	75347	79446:	gap of unknown length
	*	79447	83302:	contig of 3856 bp in length
	*	83303	83402:	gap of unknown length
	*	83403	90225:	contig of 6823 bp in length
	*	90226	90325:	gap of unknown length
	*	90326	96171:	contig of 5846 bp in length
	*	96172	96271:	gap of unknown length
	*	96272	101663:	contig of 5392 bp in length
	*	101664	101763:	gap of unknown length
	*	101764	108594:	contig of 6831 bp in length
	*	108595	108694:	gap of unknown length
	*	108695	114283:	contig of 3589 bp in length
	*	114284	114383:	gap of unknown length
	*	114384	120900:	contig of 6517 bp in length
	*	120901	121000:	gap of unknown length
	*	121001	129340:	contig of 8340 bp in length
	*	129341	129441:	gap of unknown length
	*	129441	138422:	contig of 8882 bp in length
	*	138423	138522:	gap of unknown length
	*	138523	149684:	contig of 11162 bp in length
	*	149685	149784:	gap of unknown length
	*	149785	160420:	contig of 10636 bp in length
	*	160421	160520:	gap of unknown length
	*	160521	175120:	contig of 14600 bp in length.
FEATURES				
Source				
			Location/DnaIIFiers	
			1..175120	
			/organism="Rattus norvegicus"	
			/db_xref="taxon:10116"	
			/clone="CH210-456H21"	
BASE COUNT			41850 a 41856 c41555 g 43211 t 6647 others	

DEFINITION	Homo sapiens chromosome 17 clone CTD-2020K17 map 17, *** SEQUENCEINCOMPLETE
ACCESSION	IN PROGRESS ***, 5 unordered pieces.
VERSION	AC008105
KEYWORDS	HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 116035)
TITLE	Homo sapiens chromosome 17, clone CTD-2020K17
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 116035)
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Batta,N., Beckerley,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Domela,N., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Haeos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karats,A., Leczky,J., Liu,C., Locke,K., MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (22-JUL-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	3 (bases 1 to 116035)
AUTHORS	Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Batta,N., Baatien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Goid,S., Graham,L., Grand-Pierre,N., Haeos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karats,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuppach,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Thoplam,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zaidoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (27-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Jun 27, 2002 this sequence version replaced gi:20148201. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: LB90
Center Clone name: 2020_K_17

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

	* as soon as it is available and the accession number will	
	* be preserved.	
	* 1 16830: contig of 16830 bp in length	
*	16831 16930: gap of 100 bp	
*	16931 29738: contig of 12808 bp in length	
*	29739 29838: gap of 100 bp	
*	29839 39151: contig of 3913 bp in length	
*	39152 39251: gap of 100 bp	
*	39252 45913: contig of 6662 bp in length	
*	45914 46013: gap of 100 bp	
*	46014 116035: contig of 70022 bp in length.	
FEATURES	Location/Qualifiers	
source	1. 116035	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="17"	
	/map="17"	
	/clone="CTD-2020K17"	
BASE COUNT	26107 a 30142 c 30260 g 29051 t	475 others
ORIGIN		
Alignment Scores:		
Pred. No.:	3.6e-19	116035
Score:	550.00	Matches: 133
Percent Similarity:	55.77%	Conservative: 41
Best Local Similarity:	42.63%	Mismatches: 95
Query Match:	28.80%	Indels: 43
DB:	2	Gaps: 7
US-09-972-758A-2 (1-359) x AC008105 (1-116035)		
OY	54 GlnleuglyclYArqProglYProglYgln-glySerLeuGluSerGln-----	70
Db	17680 AAGGTAGGCTTACGGCGTAGGTAAAGGCAAGGTTCCACTCAGACAGACAGCTGTCTG	17739
OY	71 -----ProProleuGlnTrpGlnAlaCyProGluSerSerCysIse	85
Db	17740 ATGCATCTTCCTTCCTCCTCCTCTTTAGACCTCTGtGCCCG-	17855
OY	85 uatrglynglylYslYsglyGlnasnGlyaspSerSerAlaGlylYaspheProPr	105
Db	17786 -----GGGAGCCCCCAAACCACCCCCTGAGCGTCATGACTCTGtGTTCCCTGCCCT	17838
OY	105 opfoalaglyValGlnProThrProGlnlaGluleuAlaGlnProcys-----	122
Db	17839 GACACCCCGGATGAGCCACTCAGAGGATGAAGAATTCTTGCGGGCTGTGCGTGCCT	17898
OY	123 ----HisaspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGluGluUtr	141
Db	17899 GGGGTGAACAGTAAGAGTCCCGGACCCAGACCCAGGGGGCTGTCAAGCGAG----	17953
OY	141 pglynglnglnglnArgGlnleuglylYslYslYslHisAtgaTargArgProSerLysLysLy	161
Db	17954 -----GCTGTGCTGGCCCGGAAAGAACCCGTGCGCGGCATCGAAGCGCA	18000
OY	161 sARGHLSTRPlysrProTyTYrLysLeuThrTrpGlnGlnLysLysPheaspGluJuy	181
Db	18001 AAGGCACTGGCGACCTCACTGAGCTGAGCTGTGGTGAAGAAACAACAGCGGATGACAG	18060
OY	181 sgInserLeuAlaAlaSerArglLeaArglaGlnmetPheAlaLysGlyGlnProValAl	201
Db	18061 GCAGGACCAAGAGGGCTCCCGGGTCCGCAAGAGATGTTCCGCAAAAGCCACCGTGGC	18120
OY	201 aPTQYrrASTTrThThGlnPheLeuMetsspaSPhtAspGlnGluGluPro-----AS	219
Db	18121 CCCCTCAACACCAACCACCTCTCTGTGATGACAGAGGCCCGGAGGAGACCCCACTTGA	18180
OY	219 pleuLYstHgylYtyrSerLysArgAlaAlaAlaLysSerAspThrSerAspas	239
Db	18181 TGTGCCCATGAGGATCTCCACCCAGGTTCCAGTGGGAGAGT-----	18223
OY	239 pasPheMetGluGlnGlyGlnGlnaspGlyLysSeraspGlyMetGlylYaspOl	259

```

Db      18224 -----GAGGCCGGGACACTGCATGGCGGGCCGACGCCA 182538
Oy      259 ySerGIuPheLLeuClArqAsPheSerGIuThrTYrGLuArqTYrHisThrGLuSerLe 279
          ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      18259 CGGTGAAGTTCCACCGAGACGACTTCTCTGTAGACTTCAGAAAGCTTCCACACGAGAGCCT 18318
Oy      279 uGlnAmMetSerLySGlnGluIleuLeuLYsgYUfYrLEuGluIleuGluLYscYseuSe 299
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      18319 GCAGGCGCCGACGACAGCAGAGCTGTGGTGCGAGACTACCTGGAGACTGTGAAGAAGCGCTGTC 18378
Oy      299 rArMeLtGuASpGUASnaSnaMrgrLeuArqLeuGluseTrLYsArGrLeuGLYglYASpas 319
          ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      18379 GCAGGCGGAGAGGAGCACTrAGAGGCGTGCACGAGCGTGCAGAGCGTGCACCGCGCAGCAgTc 18438
Oy      319 pAlArGr---ValArGrLUleuClLuLeuGluLeuASPArGrLeuArqAlaIGlASnLeuCl 338
          ||||| ||||| ||||| ||||| :||||| :||||| :||||| :||||| :||
Db      18439 CTGCCCGCAGGTGTGAGAGAGCTGGCTGTCCGAGGTCAGAGGCTCCGACCGCAAACAACAGCGC 18498
Oy      338 nLeuLeuThrGIuASnGIuDeuHISArGrIngIn 349
          ||||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      18499 GCTTCGTCAGAGACACCAGACTGTGCAGAACCGAGAG 18532

RESULT 13
LOCUS   AL662804
DEFINITION Mouse DNA sequence from clone RP23-34IC5 on chromosome 11, complete
sequence.
ACCESSION AL662804
VERSION   AL662804.17 GI:21621620
KEYWORDS  HTG.
SOURCE    house mouse.
ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (Phases 1 to 223442)
AUTHORS   Blakey,S.
TITLE     Direct Submission
JOURNAL   Submitted (26-JUN-2002) Wellcome Trust Sanger Institute, Hinton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
           humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
           On Jun 27, 2002 this sequence version replaced gi:21615602.

COMMENT   ----- Genome Center
           Center: Wellcome Trust Sanger Institute
           Genet code: SC
           Web site: http://www.sanger.ac.uk
           Contact: humquerry@sanger.ac.uk
           -----

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information from the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
constructed by the group of Piter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAce3.6.
Location/Qualifiers
1..223442
/organism="Mus musculus"

```

[illegible]

OY		181	sGlSerLeuArgIAserATgAlcArPAAGlmerPhenAlaySGlInPrOvAlA	201
Db		3993	GCAGAAGCCAGAGGCGCTTCGGGTCCGTAGAGAGATTGTTGCCAAAGSTCAGCCCCCTTGC	40523
OY		201	aProTyraNthrThrGlnPheLeuMetAspPHisaspGlnGluPurOAsPleuLy	221
Db		4053	GCCCTACAAACCACCCAGTTCATCATGATGATGCTGACCTGGAGAGGCTTAACCTC--	4110
OY		221	sThnGlyLeuTYrSerLYsArGLAlAlAlAlALysSerAspSphTrSerAspAspPph	241
Db		4111	-----GATGTGCTTCACGGGCGCTCCCA	41333
OY		241	eMeTglucLuglyGlu---GluAspGLyGYserAspGlyMetGlyGlyAspGlyse	260
Db		4134	CTTGCGCTCGCGGTGGCGAGAACGAAGCAGGAGACAGTAGTGCGCAGAGGCCACGCCCATG	41939
OY		260	rGluPheLugLnArgAspPheSerGluThrTYrGluArGTyrHisThrGlnSerLeuG1	280
Db		4194	GGAATTTCAGACAGAGGACTTCTGTGAGCGCTATGAGGCGTACACACATGAGACCTTCA	42533
OY		280	nAsmMetSerLYsGlnGluLeuIleLYsGluTYrLeuGlnGluGluGlyScLYseuSerar	300
Db		4254	GGCGCGCAGCAGAGAGAGCGTGGCCGCGAGACTCGATTGATAAGAGGCGCTGTGCACA	43133
OY		300	gMeTglAspSLysnAsnaATrGleuArTYrLeuGluSerLYsarTYrGluGluGlyGlyAspAspA1	320
Db		4314	GGCTGAGAGAGAAACTGGAGCGCTCCG-----CAGCTCCAGGGTTTCTCTCCAG	4361
OY		320	aArg-----ValArgGluLeuGlnGluLeuAspArgLeuArGLag1	335
Db		4362	CAGCAACCCCTGTCAACAGGTGGAGGAGCTGGCTGCTGAGGTGAGAGGCTCAGACATGA	4421
OY		335	uasnLeuGlnLeuLeuThrgLuasngLUeunHIsArGLngLn	349
Db		4422	A AACCAAGGGCTGGCTCAGCAGAAAGATGTGGAACCGAGAG	4464
RESULT 14				
LOCUS	BC025970			
DEFINITION	Homo sapiens, similar to putative, clone MGC:39389 IMAGE:4559410,			
ACCESSION	BC025970			
VERSION	BC025970.1			
KEYWORDS	MGC.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE	1 (bases 1 to 1238)			
JOURNAL	Strausberg/R.			
	Direct Submission			
	Submitted (18-MAR-2002) National Institutes of Health, Mammalian			
	Gene Collection (MGC), Cancer Genomics Office, National Cancer			
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
	USA			
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov			
COMMENT	Contact: MGC help desk			
	Email: gcaps-remail.nih.gov			
	Tissue Procurement: ATCC			
	cDNA Library Preparation: Rubin Laboratory			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Genome Sequence Centre,			
	BC Cancer Agency, Vancouver, BC, Canada			
	info@bcsc.bc.ca			
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Buterfield,			
	Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,			
	Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo			
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven			
	Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parveen Saedi, Jacqueline			
	Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott,			
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,			
	George Yang, Scott Zuyderduyn, Marco Marra.			

[illegible]

Db	534	AATGACAGGAGCCCGAGAGACCCACACTTGATGTGGCCCATGGGATGTCCACCCAGCT	593
OY	229	AlAlAlAlAlYsSerAspSPThSerAspAspSPHemGluGluGlyGluGlu	248
Db	594	TCGAGTGGGGAGCT-----GAG	611
OY	249	AspGlyGlySerAspGlyMetGlyYAspGlySerGluPheLeuGlnArgAspPheSer	268
Db	612	CGCGGGGACACGTATGGCGGGGCCGAGGCCAGCGCTGATGTTCCAGCGGAAAGACTTCTCT	671
OY	269	GluThrTyrgluArgTygHisThrGluSerLeuGlnAsnMetSerLySGlnGluLeuLe	288
Db	672	GAGACTTACGAACCTTCCACACCGAGAGCTTCGAGGGCCCGACACGAGAGAGCTGGTG	731
OY	289	LysGluTyrgluGluGluGlyGlyCysLeuSerArgMetGluAspGluAsnArgLeu	308
OY	732	CGAACAACCTCGAGAGCTGAGAACCGCTGTGCGAGCGGAGGAGGAGACTAGAGAGCTG	791
Db	309	ArgLeuGluSerLyArgLeuGlyGlyYAspAspAlaArg---ValArgGluLeuGlu	327
Db	792	CAGCAGCTGCAGCGCTGCACCGCGCAGCAGTCTCCCGCCAGGAGTGGAGAGACTGGCTCC	851
OY	328	GluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArg	347
Db	852	GAGTTCGAGAGCGCTCCGAGCGAAMACCGAGCGCTGCTGAGGAGAACCAAGATGTGCAC	911
OY	348	GlnGln	349
Db	912	CGAGAG	917
RESULT 15			
AK096785/c			
LOCUS	AK096785	2048 bp	mRNA
DEFINITION	Homo sapiens cDNA FLJ39466 fls, clone PROST2012353.	linear	PRI 15-JUL-2002
ACCESSION	AK096785		
VERSION	AK096785.1	GI:21756354	
KEYWORDS	Oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens prostate cDNA to mRNA, clone_1lb:PROST2 clone:PROST2012353.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.		
AUTHORS	Nishi,T., Oca,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai,Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murekawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.		
TITLE	NEO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2048)		
AUTHORS	Isogai,T. and Yamamoto,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team): 2-6-7 Kazuo-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology (RAB): cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.		
FEATURES	Location/Qualifiers		
source	1..2048		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="PROST2012353"		
	/flssue_type="prostate"		

BASE COUNT		383 a 655 c 558 g 452 t		/clone.lib="PROST2" /note="Cloning vector: pME18SFL3"	
ORIGIN					
Alignment Scores:					
Pred. NO.:		1,14e-20	Length:	2048	
Score:		545.00	Matches:	132	
Percent Similarity:		55.77%	Conservative:	42	
Best Local Similarity:		42.31%	Mismatches:	95	
Query Match:		28.53%	Indels:	43	
DB:		9	Gaps:	7	
US-09-972-758A-2 (1-359) x AK096785 (1-2048)					
OY	54	GlnLeuGlyIcylArgProGlyIleProGlyLeu-GlySerLeuGlnSerGln-----	70		
Db	1929	AAGGTAGCTTCACAGGCTAGGAGAGAGAGAAAGCTTCACCTCCACAGCAGGCTGTCTG	1870		
OY	71	-----ProProPheLeuGlnThrGlnAlaCysProGlySerCysLe	85		
Db	1869	ATGCATCCCTTCCTCCTCCTCCTCTCTTAGACCTCGTGCCCG-----	1824		
OY	85	uArgGluGlyIcyluLysGlyIleAsnGlnIAspAspSerSerAlaGlyIcylAspPhePro	105		
Db	1823	-----GGGAGGCCCCCAACACCCCTCGAGCGCTACGATCTGTGCTTCCTGCGCCT	1771		
OY	105	oProAlaGluValGlnuProThrProGluAlaGluLeuLeuAlaGlnProCys-----	122		
Db	1770	GACACCGCGGATGGAGAGCCACTCAGAGGATGAAGATCTGTGGGCTGTGGTGCGCT	1711		
OY	123	-----HisAspSerGluAlaSerIleuGlyAlaProAlaAlaGlyIcyluGlnGluTr	141		
Db	1710	GGGCTGGACAGATGAGAGCTCCCGGACCCAGAGCCAGGGGGCGTCACCGGAG-----	1656		
OY	141	pGlyGlnGlnGlnArgGlnLeuGlyIcyluLysIleAspArgArgProSerIcyluLys	161		
Db	1655	-----GCTGTGCTGGCCCCCGAAGAAACCGCTCGCGGCGCATCGAAGCCCAA	1609		
OY	161	SArgHisTrpLysProGlyTrpIcyluLysLeuThrTrpGluIcyluLysIleAspGlnIcylu	181		
Db	1608	AAGGCACTGGCGACCTCCTACCTGCAGCTGGGCTGAGAAACACAGCGGATGAGAG	1549		
OY	181	sgInSerLeuArgAlaSerArgIleArgAlaGlnuMetPheAlaLysGlnProValAl	201		
Db	1548	GCAGAGCCACAGAGGCGCTCCCGGCGCCCGAAGAGATGTTCGCCAAAGGCCAGCCGTGC	1489		
OY	201	aProTyAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnuPro-----As	219		
Db	1488	CCCCCACAAACACCCAGTCTCGATGAATGAAGGAGCCGAGAGGCCCAACTTGA	1429		
OY	219	PleuLysThrGlyLeuIcyluLysTrpSerIleArgAlaAlaIleSerAspAspThrSerAsp	239		
Db	1428	TGTGCCCCATGGGATCTCCACCCAGGTTCACGTGGGGAGAGT-----	1386		
OY	239	pAspPheMetGluGlnGlyIcyluGlnuAspGlyIcyluSerAspGlnIcyluMetGlyAspG	259		
Db	1385	-----GGGGCCGGGACAGTGAAGGGGGGGCGGCGGAGCGCA	1351		
OY	259	ySerGlnPheLeuGlnArgAspPheSerGluThrTyGlnArgTrpHisThrGlnuSerLe	279		
Db	1350	CGGTGAATTCACAGCGGAGAGACTTCTGTGAGACTTACGAACCGCTTCCACACCGAGAGCT	1291		
OY	279	uGlnAsnMetSerIcyluLysGlnIcyluLeuIcyluLysGluTrpLeuGlnuLysCysLeuSe	299		
Db	1290	GCAGGGCCGCGAAGCAGAGCTGTGCGCAAGACTTACCTGAGCTGCAGAAAGCGGCTGC	1231		
OY	299	rArgMetGlnAspGlnAsnAsnArgLeuArgLeuGlnuSerIleArgGlnuGlyIcyluAsp	319		
Db	1230	GCAGGCCGAGAGAGACTAGGAGGCTGCAGACAGCTCCAGGCGTGCACCGGCACAGACTG	1171		
OY	319	pAlaArg--ValArgGlnuGlnuGlnuLeuIleuAspArgLeuArgAlaGlnuAsnLeuG	338		

Tue Jun 17 15:12:30 2003

us-09-972-758a-2.rge

Page 18

Dd 1170 CTGCCCGCAGGTGGAGGACGTGGCTGCCGAGGTCACAGGCTCCGGAACCAAAACCAGGC 1111

Qy 338 nLeuLeuThrcLuasnGluleuHisArgGlnGln 349
::||| |||||::: :
Db 1110 GCTTCTGCAGGAAACGATGTGGAACCGAGAG 1077

Search completed: June 17, 2003, 12:08:03-
Job time : 3456 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:09:10 : Search time 72 Seconds
(without alignments)
664.402 Million cell updates/sec

Title: US-09-972-758A-2

Perfect score: 1910
Sequence: 1 MAEPPLSEYOHQPOTSNGT.....LTENELHROQERAPLSKFGD 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq.101002:*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	359	19	AAW85455
2	1910	100.0	359	23	ABP61799
3	1910	100.0	359	23	ABB76495
4	1903	99.6	359	22	AAW85465
5	308.5	16.2	152	22	AAW25726
6	237	12.4	134	22	ABG15306
7	209	10.9	349	22	ABB5958
8	155.5	8.1	281	22	ABG15303
9	147	7.7	954	22	AAU1615
10	143	7.5	288	22	ABG20362

11	141.5	7.4	1192	22	ABB59642
12	138	7.2	557	19	AAV20666
13	135.5	7.1	530	21	AAV94461
14	135.5	7.1	530	23	AAU74503
15	135.5	7.1	1749	22	ABG00839
16	135	7.1	554	23	ABB04721
17	135	7.1	555	22	AAW93869
18	134.5	7.0	510	22	ABB11764
19	134.5	7.0	510	22	AAW79741
20	134.5	7.0	546	22	AAW82808
21	133	7.0	538	21	AAW10043
22	133	7.0	648	20	AAV17946
23	133	7.0	1737	21	AAW10044
24	132.5	6.9	374	22	AAW64026
25	132.5	6.9	530	20	AAW97775
26	132.5	6.9	530	23	AAU74501
27	132.5	6.9	530	23	AAW74502
28	132	6.9	528	22	AAW78918
29	131	6.9	932	22	ABB65256
30	131	6.9	1424	22	AAW39253
31	131	6.9	1464	22	AAW41039
32	131	6.9	1898	20	AAW30795
33	130.5	6.8	819	22	AAW94316
34	130	6.8	611	20	AAV29039
35	130	6.8	611	22	AAU25510
36	129.5	6.8	1743	22	ABG10928
37	127.5	6.7	984	22	AAO12986
38	127	6.6	654	22	ABB63266
39	126.5	6.6	661	22	ABB61881
40	126.5	6.6	710	22	ABG20363
41	126.5	6.6	722	23	AAW47608
42	125.5	6.6	326	20	AAV20109
43	125.5	6.6	347	20	AAV20107
44	125.5	6.6	817	22	AAU31937
45	125	6.5	383	21	AAW13981

ALIGNMENTS

RESULT 1	AAW85455
ID	AAW85455 standard; Protein; 359 AA.
XX	XX
AC	AAW85455;
XX	XX
DT	25-FEB-1999 (first entry)
XX	XX
DE	Secreted protein encoded by clone bp783_3.
XX	XX
KW	Secreted protein; nutritional activity; immune stimulating; vaccine;
KW	tissue growth activity; haematopoiesis regulating activity;
KW	chemokine activity; haemostasis; thrombolytic activity; receptor;
KW	ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW	tumour inhibition; gene therapy.
XX	XX
OS	Homo sapiens.
XX	XX
PN	W09842739-A2.
XX	XX
PD	01-OCT-1998.
XX	XX
PF	20-MAR-1998; 98MO-US05653.
XX	XX
PR	19-MAR-1998; 98US-0044466.
XX	XX
PR	21-MAR-1997; 97US-0822167.
XX	XX
PA	(GENE) GENETICS INST INC.
XX	XX
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX	XX
PI	Racie LA, Spaulding V, Treacy M;
XX	XX

Drosophila melanog
Human neurofilament
Human p75 protein.
Human nuclear tran
Novel human diapo
Human pp2464 prote
Human polypeptide,
Human LDL binding
Human protein SFG
Human low density
MLV gag protein.
MMLV gag gene pro
MMLV gag-pol prote
Human polypeptide
Human polypeptide
Human lens epithel
Human nuclear anti
Human lens epithel
Human protein SFG
Drosophila melanog
Human polypeptide
Human polypeptide
A human trichothal
Human protein sequ
T. gondii immunoge
Novel human diapo
Novel human diapo
Drosophila melanog
Novel human diapo
Drosophila cell cy
B. burgdorferi ant
Novel human secret
Arabidopsis thaliana

DR	WT1: 1998-609890/51.
XX	
XX	N-PSDB: AAV82778.
PT	New polynucleotides encoding secreted human proteins - derived from
PT	human foetal brain, adult brain, foetal kidney, placenta or adult
PT	pitheal gland cDNA libraries.
XX	
PS	Claim 1; Page 67-68; 113pp: English.
XX	
CC	The present sequence represents a secreted protein. The polynucleotide
CC	and secreted protein are predicted to have biological activities which
CC	could make them suitable for treating, preventing or ameliorating medical
CC	conditions in humans and animals, although no supporting data is given.
CC	Suggested activities include nutritional activity, immune stimulating
CC	(e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC	activity, tissue growth activity, activin/inhibin activity,
CC	chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC	receptor/ligand activity, anti-inflammatory activity, cachectin/tumour
CC	invasion suppressor activity, and tumour inhibition activity (no data is
CC	given in the specification to support these activities). The
CC	polynucleotide is also stated to be useful for gene therapy.
XX	
SO	Sequence 359 AA:
	Query Match 100.0%; Score 1910; DB 19; Length 359;
	Best Local Similarity 100.0%; Pred. No. 5.9e-153;
	Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MAEPFLESYOHQPTNSCTGAAAVOEELNPERPGAERVRVEDSRMSRAFPOLGARG 60
Db	1 MAEPFLESYOHQPTNSCTGAAAVOEELNPERPGAERVRVEDSRMSRAFPOLGARG 60
QY	61 PEGGSELSQPPPLQTOACPESSCLRGKNGQNDSSAGSGDFPPAVERPTPEALLAQ 120
Db	61 PEGGSELSQPPPLQTOACPESSCLRGKNGQNDSSAGSGDFPPAVERPTPEALLAQ 120
QY	121 PCHDSKSKUGAPAAAGGEEEMGQOQROLGKKKRRRPSKKRRHMKPYIKLTWEKKKFE 180
Db	121 PCHDSKSKUGAPAAAGGEEEMGQOQROLGKKKRRRPSKKRRHMKPYIKLTWEKKKFE 180
QY	181 KQSLRARIRAREMPAKGOPYVRYTTFGLMDHDOEEDLKTGLYSKRAAKSDTSDDD 240
Db	181 KQSLRARIRAREMPAKGOPYVRYTTFGLMDHDOEEDLKTGLYSKRAAKSDTSDDD 240
QY	241 FMEGGEGEDGSDGSDGSEFLORDFSEYERYHTESLQMSKOEILKEYLEDEKLSR 300
Db	241 FMEGGEGEDGSDGSDGSEFLORDFSEYERYHTESLQMSKOEILKEYLEDEKLSR 300
QY	301 MEDENNRLRLSEKRLGGDDARVRELELELDRLARENOLLTENELHQOEARPLSKTGD 359
Db	301 MEDENNRLRLSEKRLGGDDARVRELELELDRLARENOLLTENELHQOEARPLSKTGD 359
RESULT 2	
ABP61799	
ID	ABP61799 standard; Protein: 359 AA.
XX	
AC	ABP61799;
XX	
DT	04-OCN-2002 (first entry)
DE	Human polypeptide SEQ ID NO 153.
XX	
KW	Human: cytosolic; antirheumatic; antiarthritic; vulnery; analgesic;
KW	antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW	neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW	antitumor; fungicide; antidiabetic; antiashtmatic; antiallergic;
KW	immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW	cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW	stem cell; growth factor; nervous system disease; neuropathy;
KW	Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW	osteoporosis; severe combined immunodeficiency; SCID; infection;
KW	multiple sclerosis; rheumatoid arthritis; gene therapy.

OS Homo sapiens.
PN US2002065394-A1.
PP 30-MAY-2002.
PR 22-DEC-2000; 2000US-0745763.
XX 18-MAR-1998; 98US-0040963.
XX (JACO/) JACOBS K.
XX (MCCO/) MCCOY J M.
XX (LAVALLIE) LAVALLIE E R.
XX (COLL/) COLLINS-RACIE L A.
XX (EVAN/) EVANS C.
XX (MERB/) MERBERG D.
XX (TREAC/) TREACY M.
XX (SPAUD/) SPAUDING V.
PI Jacobs K, MCCOY JM, LAVALLIE ER, COLLINS-RACIE LA, EVANS C;
PI Meireis D, Treacy M, Spaulding V;
XX WPI: 2002-582343/62.
XX DR N-PSDB: ABQ92015.
XX Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
XX
XX Claim 44; Page 112-113; 284pp; English.
XX
XX The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (CDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, anti-inflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC activities. (I) can be used to manipulate stem cells in culture to give
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and
CC regeneration of nerve and brain tissue and is useful for the treatment of
CC central and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC In various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial and fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polypeptide of the invention.
XX
XX Sequence 359 AA;
SQ
Query Match 100.0%; Score 1910; DB 23; Length 359;
Best Local Similarity 100.0%; Pred. No. 5.9e-153;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match	Best Local Similarity	99.6%	Score 1903:	DB 22:	Length 359:				
Matches	358:	Conservative	0:	Mismatches	1:	Indels	0:	Gaps	0:
QY	1	MAEFLSEYQIOPOTS	NCCTGA	AAVQELINPERP	PGAEIRVEEDSR	WMSRAFPOLGSGRPG	60		
Db	1	MAEFLSEYQIOPOTS	NCCTGA	AAVQELINPERP	PGAEIRVEEDSR	WMSRAFPOLGSGRPG	60		
QY	61	PEGGSLSLSEOPRPL	LOTQACPESSSCL	RESEKQNGDSS	AGGDFPPPAVEP	ETPPAEELIAQ	120		
Db	61	PEGGSLSLSEOPRPL	LOTQACPESSSCL	RESEKQNGDSS	AGGDFPPPAVEP	ETPPAEELIAQ	120		
QY	121	PCHDSSEASKLAP	PAAGSEENGGOO	ROLGKKHRRP	PKKRRHMPYK	LTLWEKKKFFDE	180		
Db	121	PCHDSSEASKLAP	PAAGSEENGGOO	ROLGKKHRRP	PKKRRHMPYK	LTLWEKKKFFDE	180		
QY	181	KOSTIRASTIRAE	MFAGOPVAPYNT	QFGLMDHDOE	EBPDLTGILSK	RAAKSDTSDDD	240		
Db	181	KOSTIRASTIRAE	MFAGOPVAPYNT	QFGLMDHDOE	EBPDLTGILSK	RAAKSDTSDDD	240		
QY	241	FMEGGGEEDGSD	GMGGSEFLORDP	SEYTERHTESL	OWMSKOELIK	ETLELEKCLR	300		
Db	241	FMEGGGEEDGSD	GMGGSEFLORDP	SEYTERHTESL	OWMSKOELIK	ETLELEKCLR	300		
QY	301	MEDENNRLRL	ESKRLGGDDAV	RELELELDRL	AEINQLTL	ENELHROE	RAPLSKFGD	359	
Db	301	MEDENNRLRL	ESKRLGGDDAV	RELELELDRL	AEINQLTL	ENELHROE	RAPLSKFGD	359	

ID	AA025726	standard; Protein; 152 AA.
AC	AA025726:	
XX		
DT	16-OCT-2001	(first entry)
DE	Human protein sequence SEQ ID NO:1241.	
XX		
KW	Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antiarthritis; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; vitruide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antiagregant; haemostatic; vulnery; anticler; osteoptich; eczema; dermatological; antiallergic; antistatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antiatherosclerit; Rheumatoid arthritis; septic shock; pancreatitis; antiatherosclerit; Rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; genetic dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; cardiac dysfunction; haematopoietic disorders; osteoporosis; severe combined immunodeficiency; eczema; allergic rhinitis; asthma; diabetes; cancer; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative and neurological disorders.	
XX		
Sequence	152 AA:	
Query Match	16.2%;	Score 308.5; DB 22; Length 152;

	CC	imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity.
	CC	The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and amino acid sequences ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.
	CC	Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
SQ	Sequence	281 AA;
	Query Match	8.1%; Score 155.5; DB 22; Length 281:
	Blast Local Similarity	26.0%, Pred. No. 6e-05;
MATCHES	60: Conservative	24: Mismatches 86: Indels 61: Gaps 122
OY	14 QTSNCTG-AAVVOELINPERPPGAEEVPPEE-----DSRMOSAFPOLGGPGR-PGE	64
Db	67 RASACEDGPITLLRPSHSCHPRGAEERYPSTACLORSMEGNR-NALDOVGSLKSPWIGE	125
OY	65 GSLESOPP-----LQTQACESSCLREGEKGQGDDSSAGDPPEPAEV----EPPIE	114
Db	126 YKEBRHPNPQNANYHQTLACPARECFRGOG-----PGLRRQAEP-L-S	170
OY	115 AELLAPCPHDESEASKLGAPAAGE-----EEWGQQQRQLGKKKHRRRPSSKKKR--H	163
Db	171 SSSLSSHOSPQSOTILPLXLRRLSGCGNPYLCPWSOPLPETTAHGSKGPSRSNCSS	230
OY	164 WKPYPKLT-----WEKKRFDEKOSLRASRIR-----AENFAKGP	199
Db	231 WKNMVSSTACPSHEFLMRSRKFERSASTASTPCRMFTVARTHQRIPLKARCTP	281
 RESULT 9 AAU14615 ID AU14615 standard; Protein: 954 AA. XX AAU14615: DT 24-Oct-2001 (first entry) DE Novel bone marrow polypeptide #14. KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic; KM haematopoiesis; myeloid; lymph cell disorder; tissue regeneration; KW wound healing; nutritional supplement; immune disorder; KW severe combined immunodeficiency; SCID. XX Homo sapiens. OS WO200157187-A2. PN 09-Aug-2001. PF 05-Feb-2001; 2001WO-USO3782. XX 03-Feb-2000; 2000US-0496914. PR 20-JUN-2000; 2000US-0598075. PR 19-JUL-2000; 2000US-0620325. PR 30-Nov-2000; 2000US-0250683. XX PA (HYSEQ-) HYSEQ INC. PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ, PI Ren F, Drmanac RT; DR N-PSDB; AAS22920. XX WTPI: 2001-488875/53. DR TT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and gene therapy - XX		

XX Claim 10; Page 215-217; 392pp; English.

PS ANU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify modulators of their expression and activity.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).

XX Sequence 954 AA:

Query Match 7.7%; Score 147; DB 22; Length 954;
Best Local Similarity 21.6%; Pred. No. 0.0015;

Matches 97; Conservative 65; Mismatches 146; Indels 142; Gaps 19;

QY 25 QBELNPPPPGAEER-----VEPDSRWOS--RAF-----POLGRRPGPEG----- 64
DB 189 QDELQSKOSKGLLEEYHRLSPPLPHEE--RAQSPRLATEEPPQ-----GPEGPEWKE 242
QY 65 -----GSLSEPPPIQTOA--CPRESSCLREGKQNGDSSAGSGFPPPAE-----VE 110
DB 243 AEELGDSASLSLQSLQREDAVSPAPACEKGKQHSQAEELGPGDEADDEKAVAS 302
QY 111 PTP-----BAELLAQPCDSEASKLGAAPAGEEEMGQOOROLGKKR----- 152
DB 303 PTPVSPPEVRSTEPYAPPEQLSEAA--LKAMEAAVQVLEQQRHLSEKQEMQGLRKL 361
QY 153 -----HRRPSSKKRHKMPYKLTWEKKKPFQKOSLSRASRIRAEMFAGQPYA 201
DB 362 QDEEEETLRLHQKEQSLRETLQALIEEERARMEESORLSWLRAGVOSTQA-- 419
QY 202 PYNTTQFLMDHDEEPLDKTGLYSKRAAKSD-----DTSDDFMEEGEGEDGS- 252
DB 420 -----DEDOIARAQDEASLQKLREELSQOKAEVASLQFQKNRQMLQDEKETENSE 469
QY 253 -----DGMGDGSSEFLQRFSEYERY-----HTE-----SLQ 280
DB 470 KSEQALNAKAKKALQQLREQLGERKPAVATLEKEHSAELRLCSSLEAKHREVVSLSQ 529
QY 281 NMSQQLKEYLELEKCLSRMEDENNRLKESKRLGSGDARVREL-----ELE 328
DB 530 KKIDEAQKEBAQLQKQGYE---HRVHOKSYHVAQYEHLSLSLREKROEVEGEHRR 586
QY 329 IDRLRAENLQILIT-----ENELHROERAPL 354
DB 587 LDKMKHEHQYMAKAREQYEAERKQRAEL 616

RESULT 10
ID ABG20362 standard: Protein: 288 AA.
XX AC ABG20362;
XX

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #20353.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;

KW Food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS84549.

XX

XX

XX

PS

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Claim 20; SEQ ID No 50721; 103pp; English.

The invention relates to isolated polynucleotide (I) and

polypeptide (II) sequences. (I) is useful as hybridisation probes,

polymerase chain reaction (PCR) primers, oligomers, and for chromosome

and gene mapping, and in recombinant production of (II). The

polynucleotides are also used in diagnostics as expressed sequence tags

for identifying expressed genes. (I) is useful in gene therapy techniques

to restore normal activity of (II) or to treat disease states involving

(II). (II) is useful for generating antibodies against it, detecting or

quantitating a polypeptide in tissue, as molecular weight markers and as

a food supplement. (II) and its binding partners are useful in medical

imaging of sites expressing (II). (I) and (II) are useful for treating

disorders involving aberrant protein expression or biological activity.

The polypeptide and polynucleotide sequences have applications in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits to assess biodiversity

and to produce other types of data and products dependent on DNA and

amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 288 AA:

Query Match 7.5%; Score 143; DB 22; Length 288;
Best Local Similarity 27.5%; Pred. No. 0.0007;

Matches 69; Conservative 24; Mismatches 64; Indels 94; Gaps 14;

QY 50 RAPQUL---GGRPG-----PEGESLSQPPPIQTOACPPSSCLREGKQGNDDDS 97
DB 9 RAGVQLNQGARGPGSTAGDCCRRTCTL-----ALRLRA-----RSEKRODSALER 56
QY 98 SAGGDFPPPAVEPTPEAELLAPCHDSEASKLGAAPAGE-----EWM-GQOOR----- 146
DB 57 S--GRAPPPL-----LRGLGANGAGARGGCGCDPRWGCQWPRGRF 95
QY 147 -----QLGKKKHHRRPSKKKHKMPYKLTW-----PEKKKFPDK----- 181
DB 96 EIKKKLTKAKKKEKKKEDEKKKLLQIOESOVTSNNKRRSRKRDLDKKSOAM 155
QY 182 QSLRASRIR-----AEMFAKQPVAPYNTTQFLMDHDHDEEPLDKTGLYSKRAAKSDDT 236

[illegible]

Db	349--SQEVEVSQPEVEAQPEVEPOSEVSESQ-----PEABSHSPE---TQAEVE	390
Qy	177 KFDKQKSI--RASIIRAEFMFAKGPVAPYNTTQFLMDHDQDEPDLTKGLYSKRAAKSDT	236
Db	391 AQPEVESLPEAESQPE--AESQP-----EREVEE---AEKISDNEVPT	429
Qy	237 SDDPEEGCGEFGDGGSDGCGGSEFLQDFSETERYHTESLQNNKSKOLEIKKEYLEBK	296
Db	430 TEASIMELTVE--GIEDGITAAAMDNIYPEELAEASDKQTE--LESDDQSPYTALE-EQ	485
Qy	297 CLSRMDENNRLRLSKRLGDDARVRELELELDRLRAENLQ--LTENELHQOERAP	353
Db	486 AVPEIEQEK---PPEQITLAD-----ETEQDSQAQSPNEBPVEIAPQGHTEATAP	534
RESULT 12		
AAV20666		
ID	AAV20666 standard; Protein: 557 AA.	
XX	AAV20666;	
AC		
XX	22-JUL-1999 (first entry)	
DT		
XX	Human neurofilament-M wild type protein fragment 8.	
DE		
XX	Human: beta-amyloid precursor protein; beta-APP; diagnosis: cancer;	
KW	frameshift mutation; age-related disease; neurodegenerative disorder;	
KW	Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;	
KW	Huntington's disease; multiple sclerosis; alcoholic liver disease;	
KW	diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;	
KW	ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;	
KW	neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;	
KW	glial fibrillary acidic protein; GFAP; p53; semaphorin IIR; HUPF-1;	
KW	bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;	
KW	high mobility group protein-C; neuroendocrine specific protein A.	
XX		
OS	Homo sapiens.	
XX		
PN	W09845322-A2.	
XX		
PD	15-OCT-1998.	
XX		
PP	02-APR-1998; 98WO-1B00705.	
PR	10-APR-1997; 97US-0043163.	
XX		
XX	(UYUT-) RIJXSUNIV UTRECHT.	
PA	(ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.	
PA	(UYRO-) UNIV ROTTERDAM ERASMUS.	
XX		
PL	Butbach JPH, Grosveld FG, Van Leeuwen FW;	
DR	WPI; 1998-609901/51.	
XX	N-PSDB; AAX75759.	
PT	Diagnosing disease by detecting frameshift mutations in RNA or	
PT	corresponding protein mutations - used to diagnose cancer and	
PT	neurological diseases, particularly Alzheimer's disease, and also	
PT	for treatment and prevention with specific ribozymes or wild-type	
PT	RNA	
XX		
PS	Disclosure: Figure 8; 258pp; English.	
XX		
CC	This invention describes a novel method for the diagnosis of a disease	
CC	caused by, or associated with, an RNA molecule that has a frameshift	
CC	mutation. The method is used to diagnose age-related diseases, especially	
CC	cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's	
CC	disease, Down's syndrome, myotonic dystrophy, Huntington's disease,	
CC	multiple sclerosis, alcoholic liver disease, diabetes mellitus type II	
CC	and many others listed) or susceptibility to these disorders. The method	
CC	allows a definitive diagnosis of Alzheimer's disease in living patients,	
CC	at an early stage. It is based on the observation that disease may be	
CC	caused by mutations in RNA rather than DNA. The invention describes the	

FT /note= "Encoded by GFA"
 XX WO200190751-A1.
 XX PD 29-NOV-2001.
 XX PF 23-MAY-2001; 2001WO-US16639.
 XX PR 23-MAY-2000; 2000US-207141P.
 XX PA (SCLRI) SCRIPPS RES INST.
 XX PI Tan EM, Ochs RL, Chan EKL, Muro Y;
 XX DR WPI; 2002-122026/16.
 XX DR N-PSDB; AAS20483.
 XX PT Screening a patient for atopic dermatitis, comprises determining the
 PT presence of antibodies against nuclear transcription co-activator p75
 PT in the sera of the patient -
 XX PS Disclosure; Page -: 22pp; English.
 XX CC The invention relates to a method of screening a patient for atopic
 CC dermatitis, comprising determining the presence of antibodies against
 CC nuclear transcription co-activator p75, in sera of the patient. The
 CC method is used for diagnosing patients with atopic dermatitis, and for
 CC detecting atopic dermatitis in patients suffering from other conditions
 CC such as asthma or interstitial cystitis. This sequence represents the
 CC human p75 protein.
 XX SQ Sequence 530 AA:
 Query Match 7.1%; Score 135.5; DB 23; Length 530;
 Best Local Similarity 21.8%; Pred. No. 0.0067;
 Matches 79; Conservative 59; Mismatches 110; Indels 115; Gaps 18;
 QY 7 SEYQHOPOTSNGTAAYOEELNPRPPGAERVPEDSRMOSRAFPQ-----L 55
 DB 93 SSOQATAKOSNASDSVEVEEK-----ETSVSKEDTDHEKASNEDEVTKAVDITPK 143
 QY 56 GGRPG-PEGESLSEOPRPIOTOCPESSCLREGEKONGDDSSAGDPPPAFV---E 110
 DB 144 AARRGRKKRAEKQVETTERAGVTTTATASVNLKVSFKRG-----PAATEVKIPK 192
 QY 111 PTPPEALLAOPCHDEASKLGAAPAGGEEMWGOQOLGKKKRRRPSKKRHHMKPYVKL 170
 DB 193 PGRGRKWKVQPC-PESDII-----TEEDSKSKKKGGGK-----QPKK----- 229
 QY 171 TWEKKKKFDEKQSLASRIKRAEMFAKQGVAPYNTTQFLMDHDQEEFDLKTG---LYSK 227
 DB 230 ---QPKKDEEOK-----EEDKPKRPEDKKREKKKEVESK 260
 QY 228 RA-AKSDTSDDDMEEGEGEDGGSDGMGGDGEFFLQDPSFETYEY-----YHTESLQ 280
 DB 261 RKNLAKTGVSTSDSEEGDDOEGEKKRKG-----RNFQTAHRRMLKGMQHEKEAD 313
 QY 291 NMSKOELIKYLELEKLSRMEDENNR---LRLESKRUGGDARVRELELEL-DRLRAEN 336
 DB 314 RKRQGE---EQWETBQ---QNKDEGKKPEVKKVEKKRFTSMDSRLQRIHAELKNSLKIDN 367
 QY 337 LQL 339
 DB 368 LDV 370
 RESULT 15
 ABG00839 standard; Protein: 1749 AA.
 ID ABG00839;
 AC ABG00839;
 XX 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #830.
 XX KW Human: Chromosome mapping; gene mapping; gene therapy; forensic;
 XX OS food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HSE-) HXSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS65026.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX PS Claim 20; SEQ ID NO 31198; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX SQ Sequence 1749 AA:
 Query Match 7.1%; Score 135.5; DB 22; Length 1749;
 Best Local Similarity 21.9%; Pred. No. 0.032;
 Matches 94; Conservative 53; Mismatches 123; Indels 159; Gaps 19;
 QY 20 GAAVVOEELNPRPPGAERVPEDSRMOSRAFPOLGRRPPEGESLSEOPPLQIQC 79
 DB 160 GAA---KOSNPKSSBGOPEAGPEGAQERPSQAAPAV-----EAEQGSSQAP-----RK 205
 QY 80 PESSCLREGEKONGD-----SSAGDPPPAEVEPTPEALLAOP 121
 DB 206 PEGAQARTAGSGALRDVSEELSRQLEDLTSTYCDNNOGGGEGGAGGP-----AEP 258
 QY 122 CHDSEASKL-----GAPAGGEEFWG-----QQQROLGKKKRRRPSKKRRH 163
 DB 259 -EDAERSTYVARNGEPEPTPVNGEKKPSKGDPTETRIOSDDEVGDHRHRRPQEKKAK 317
 QY 164 W-----KPYKLTWEKKKKFDEKQSLASRIKRAEMFAK 197

THIS PAGE BLANK (USPTO)